

150.5
J812
v.5

JUL 1 1971

journal of applied psychology

KENNETH E. CLARK, Editor
University of Rochester

Consulting Editors

GEORGE E. BRIGGS, *Ohio State University*

MARVIN D. DUNNETTE, *University of Minnesota*

NORMAN FREDERIKSEN, *Educational Testing Service*

LEONARD D. GOODSTEIN, *University of Cincinnati*

EDWIN R. HENRY, *University of Rochester*

JOHN HOLLAND, *American College Testing Program*

CLIFFORD E. JURGENSEN, *Minneapolis Gas Company*

CLARK L. WILSON, *Instituto Centroamericano de Administracion de Empresas*

DONALD G. MARQUIS, *Massachusetts Institute of Technology*

LAURENCE S. MCGAUGHRAN, *University of Houston*

QUINN MCNEMAR, *University of Texas*

WARREN T. NORMAN, *University of Michigan*

HAROLD F. ROTHE, *Beloit Corporation*

THOMAS A. RYAN, *Cornell University*

ALEXANDER G. WESMAN, *Psychological Corporation*

Volume 53, 1969

HELEN ORR
Managing Editor

CLAIRE PILLIOD
Technical Editor

HENRY S. SALMON
Advertising and
Business Manager

ROBERT D. GORMAN
Subscription Manager

Published bimonthly by the American Psychological Association, Inc.
Prince and Lemon Sts., Lancaster, Pa. 17604 and 1200 Seventeenth St., N.W.
Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.

© 1969 by the American Psychological Association, Inc.

Contents of Volume 53

| | |
|---|-----|
| ABRAHAM, NORMAN M. Off-Quadrant Comment..... | 66 |
| ALF, EDWARD F. See CURTIS, ERVIN W. | |
| ALLEN, JOHN CHARLES. Multidimensional Analysis of Worker-Oriented and Job-Oriented Verbs... | 73 |
| ALLOWAY, THOMAS M. See BERNSTEIN, DOUGLAS A. | |
| ANASTASI, ANNE, & SCHAEFER, CHARLES E. Biographical Correlates of Artistic and Literary Creativity in Adolescent Girls..... | 267 |
| ANDERSON, LYNN R. See COLLAROS, PANAYIOTA A. | |
| APPEL, VALENTINE, & FEINBERG, M. R. Recruiting Door-to-Door Salesmen by Mail..... | 362 |
| ARONSON, ELLIOTT. See LANDY, DAVID. | |
| ATHELSTAN, GARY T. See STONE, THOMAS K. | |
| ATTWOOD, DENNIS A., & WIENER, EARL L. Automated Instruction for Vigilance Training..... | 218 |
| BARRELL, R. P. See RHINEHART, J. B. | |
| BARRETT, GERALD V., THORNTON, CARL L., & CABE, PATRICK A. Relation between Embedded Figures Test Performance and Simulator Behavior..... | 253 |
| BENSON, PURNELL H. Marginal Productivity Procedure for Staff Selection..... | 124 |
| BERDIE, RALPH F. Consistency and Generalizability of Intraindividual Variability..... | 35 |
| BERGUM, BRUCE O., & DOOLEY, ROGER P. Forced-Choice Preferences and Response Latency..... | 396 |
| BERNSTEIN, DOUGLAS A., & ALLOWAY, THOMAS M. On the Use of Alternative Organisms..... | 506 |
| BERSH, PHILIP J. A Validation Study of Polygraph Examiner Judgments..... | 399 |
| BHATNAGER, R. P. A Study of Some EPPS Variables as Factors of Academic Achievement..... | 107 |
| BLAKE, RICHARD. Comparative Reliability of Picture Form and Verbal Form Interest Inventories... | 42 |
| BLOOD, MILTON R. Work Values and Job Satisfaction..... | 456 |
| BOLTON, DALE L., & HICKEY, MICHAEL E. Effect of Interviews on Teacher Selection Decisions... | 501 |
| BOHR, RONALD H., & SWERTLOFF, ARNOLD B. Work Shift, Occupational Status, and the Perception of Job Prestige..... | 227 |
| BOHRNSTEDT, GEORGE W. See FORD, ROBERT N. | |
| BORGATTA, EDGAR F. See FORD, ROBERT N. | |
| BOUCHARD, THOMAS J., JR. See Monograph listing. | |
| BRAY, DOUGLAS W. See GRANT, DONALD L. | |
| BROWN, I. D., TICKNER, A. H., & SIMMONDS, D. C. V. Interference between Concurrent Tasks of Driving and Telephoning..... | 419 |
| BRUNNER, G. ALLEN. See PAINE, FRANK T. | |
| BRUVOLD, WILLIAM H., & GAFFEY, WILLIAM R. Rated Acceptability of Mineral Taste in Water: II. Combinatorial Effects of Ions on Quality and Action Tendency Ratings..... | 317 |
| BRYAN, JUDITH F. See LOCKE, EDWIN A. | |
| CABE, PATRICK A. See BARRETT, GERALD V. | |
| CARROLL, STEPHEN J., & TAYLOR, WILLIAM H. JR. Validity of Estimates by Clerical Personnel of Job Time Proportions..... | 164 |
| COLLAROS, PANAYIOTA A., & ANDERSON, LYNN R. Effect of Perceived Expertness upon Creativity of Members of Brainstorming Groups..... | 159 |
| COSENTINO, JOSEPH. See KIPNIS, DAVID. | |
| CROPLEY, A. J., & FIELD, T. W. Achievement in Science and Intellectual Style..... | 132 |
| CURTIS, ERVIN W., & ALF, EDWARD F. Validity, Predictive Efficiency, and Practical Significance of Selection Tests..... | 327 |
| DENELSKY, GARLAND Y., & MCKEE, MICHAEL G. Prediction of Job Performance from Assessment Reports: Use of a Modified Q-Sort Technique to Expand Predictor and Criterion Variance..... | 439 |
| DEWOLFE, A. S. See RHINEHART, J. B. | |
| DICKINSON, TERRY L. See NAYLOR, JAMES C. | |
| DOLL, RICHARD E., GUNDERSON, E. K. ERIC. Occupational Group as a Moderator of the Job Satisfaction-Job Performance Relationship..... | 359 |
| DOOLEY, ROGER P. See BERGUM, BRUCE O. | |
| DUBIN, JERRY, A., OSBURN, HOBART, & WINICK, DARVIN M. Speed and Practice: Effects on Negro and White Test Performances..... | 19 |
| EISENMAN, RUSSELL. Creativity and Academic Major: Business versus English Majors..... | 392 |
| EVANS, MARTIN G. Conceptual and Operational Problems in the Measurement of Various Aspects of Job Satisfaction..... | 93 |
| EVANS, MARTIN G. Convergent and Discriminant Validities between the Cornell Job Descriptive Index and a Measure of Goal Attainment..... | 102 |
| EWEN, ROBERT B. The GRE Psychology Test as an Unobtrusive Measure of Motivation..... | 383 |
| FARNSWORTH, KIRK E. Vocational Interests of Women: A Factor Analysis of the Women's Form of the SVIB..... | 353 |
| FARRIS, GEORGE F. Organizational Factors and Individual Performance: A Longitudinal Study..... | 87 |
| FARRIS, GEORGE F., & LIM, FRANCIS G., JR. Effects of Performance on Leadership, Cohesiveness, Influence, Satisfaction, and Subsequent Performance..... | 490 |
| FEINBERG, M. R. See APPEL, VALENTINE. | |

| | | |
|---|--|-----|
| FIELD, T. W. | See CROPLEY, A. J. | |
| FORD, ROBERT N., BORGATTA, EDGAR F., & BOHRNSTEDT, GEORGE W. | Use of the Work Components Study with New College-Level Employees..... | 367 |
| FREEBERG, NORMAN E. | Relevance of Rater-Ratee Acquaintance in the Validity and Reliability of Ratings..... | 518 |
| FURCON, J. | See GOODMAN, P. | |
| GAFFEY, WILLIAM R. | See BRUVOLD, WILLIAM H. | |
| GETTYS, CHARLES F. | See MARTIN, DAVID W. | |
| GOLDMAN, MORTON. | See O'LEARY, CHARLES. | |
| GOLDSTEIN, IRWIN L., JOHNSTON, WILLIAM A., & HOWELL, WILLIAM C. | Complex Vigilance: Relevant and Irrelevant Signals..... | 45 |
| GOODMAN, P. FURCON, J., & ROSE, J. | Examination of Some Measures of Creative Ability by the Multitrait-Multimethod Matrix..... | 240 |
| GOSS, ALLEN. | Predicting Vocational Success for Neuropsychiatric Patients with the Edwards Personal Preference Schedule..... | 250 |
| GRAEN, GEORGE. | See Monograph listing. | |
| GRANT, DONALD L., & BRAY, DOUGLAS W. | Contributions of the Interview to Assessment of Management Potential..... | 24 |
| GRIFFIN, J. E. | See RHINEHART, J. B. | |
| GUNDERSON, E. K. ERIC. | See DOLL, RICHARD E. | |
| HACKMAN, J. RICHARD. | See LAWLER, EDWARD E., III. | |
| HALL, WALLACE B., & MACKINNON, DONALD W. | Personality Inventory Correlates of Creativity among Architects..... | 322 |
| HARMON, LENORE W. | Predictive Power over Ten Years of Measured Social Service and Scientific Interests among College Women..... | 193 |
| HENEMAN, HERBERT G., III. | See SCHWAB, DONALD P. | |
| HICKEY, MICHAEL E. | See BOLTON, DALE L. | |
| HILL, WALTER. | A Situational Approach to Leadership Effectiveness..... | 513 |
| HILLE, STANLEY J. | See PAINE, FRANK T. | |
| HINRICH, J. R. | Comparison of "Real Life" Assessments of Management Potential with Situational Exercises, Paper-and-Pencil Ability Test and Personality Inventories..... | 425 |
| HINRICH, J. R. | Correlates of Employee Evaluations of Pay Increases..... | 481 |
| HOWELL, WILLIAM C. | See GOLDSTEIN, IRWIN L. | |
| HULIN, CHARLES J. | Sources of Variation in Job and Life Satisfaction: The Role of Community and Job-Related Variables..... | 279 |
| HUNDAL, P. S. | Knowledge of Performance as an Incentive in Repetitive Industrial Work..... | 224 |
| IVANCEVICH, JOHN M. | Perceived Need Satisfactions of Domestic versus Overseas Managers..... | 274 |
| JACKSON, RUSSELL H. | See WIENS, ARTHUR N. | |
| JOHNSON, RICHARD W. | Effectiveness of SVIB Academic Interest Scales in Predicting College Achievement..... | 309 |
| JOHNSTON, WILLIAM A. | See GOLDSTEIN, IRWIN L. | |
| KANUNGO, RABINDRA N. | Brand Awareness: Differential Roles of Fittingness and Meaningfulness of Brand Names..... | 140 |
| KIPNIS, DAVID, & COSENTINO, JOSEPH. | Use of Leadership Powers in Industry..... | 460 |
| KORMAN, ABRAHAM K. | Self-Esteem as a Moderator in Vocational Choice: Replications and Extensions..... | 188 |
| KUNERT, KENNETH M. | Psychological Concomitants and Determinants of Vocational Choice..... | 152 |
| LANDY, DAVID, MCCUE, KATHLEEN, & ARONSON, ELLIOTT. | Beyond Parkinson's Law: III. Effect of Protractive and Contractive Distractions on the Wasting of Time on Subsequent Tasks..... | 236 |
| LAWLER, EDWARD E., III, & HACKMAN, J. RICHARD. | Impact of Employee Participation in the Development of Pay Incentive Plans: A Field Experiment..... | 467 |
| LEBO, DELL. | See LONG, THOMAS REX. | |
| LEMELY, J. H. | See WRIGHT, JACK M. | |
| LIM, FRANCIS G., JR. | See FARRIS, GEORGE F. | |
| LOCKE, EDWIN A., & BRYAN, JUDITH F. | Knowledge of Score and Goal Level as Determinants of Work Rate..... | 59 |
| LONG, THOMAS REX, LEBO, DELL, & OGDEN, DOROTHY. | A Simplified Method for Increasing Mail-Sorting Efficiency..... | 510 |
| LUNNEBORG, CLIFFORD E., & LUNNEBORG, PATRICIA W. | Architecture School Performance Predicted from ASAT, Intellectual, and Nonintellectual Measures..... | 209 |
| LUNNEBORG, PATRICIA W. | See LUNNEBORG, CLIFFORD E. | |
| MACKINNON, DONALD W. | See HALL, WALLACE B. | |
| MANAUGH, THOMAS S. | See WIENS, ARTHUR N. | |
| MARKS, EDMUND, & WEBB, SAM C. | Vocational Choice and Professional Experience as Factors in Occupational Image..... | 292 |
| MARTIN, DAVID W., & GETTYS, CHARLES F. | Feedback and Response Mode in Performing a Bayesian Decision Task..... | 413 |
| MATARAZZO, JOSEPH D. | See WIENS, ARTHUR N. | |
| MCCLELLAND, JAMES N., & RHODES, FEN. | Prediction of Job Success for Hospital Aides and Orderlies from MMPI Scores and Personal History Data..... | 49 |

| | |
|---|-----|
| McCUE, KATHLEEN. See LANDY, DAVID. | |
| McKEACHIE, W. J. Interaction of Achievement Cues and Facilitating Anxiety in the Achievement of Women | 147 |
| McKEE, MICHAEL G. See DENELSKY, GARLAND Y. | |
| McNAMARA, W. J. See WOLLOWICK, HERBERT B. | |
| McNEMAR, QUINN. Moderation of a Moderator Technique | 69 |
| MEGARGEE, EDWIN I. Influence of Sex Roles on the Manifestation of Leadership | 377 |
| MOLLENKOPF, WILLIAM G. Some Results of Three Basic Skills Training Programs in an Industrial Setting | 343 |
| NASH, ALLAN N. See PAINE, FRANK T. | |
| NAYLOR, JAMES C., & DICKINSON, TERRY L. Task Structure, Work Structure, and Team Performance | 167 |
| NORMAN, WARREN T. See PASSINI, FRANK T. | |
| O'BRIEN, GORDON E., & OWENS, A. G. Effects of Organizational Structure Correlations between Member Abilities and Group Productivity | 525 |
| OGDEN, DOROTHY. See LONG, THOMAS REX | |
| O'LEARY, CHARLES J., & GOLDMAN, MORTON. Comparison of Several Patterns of Communication ... | 451 |
| O'NEILL, PAUL J. See TANOFKY, ROBERT. | |
| OSBURN, HOBART. See DUBIN, JERRY A. | |
| OWENS, A. G. See O'BRIEN, GORDON E. | |
| OWENS, W. A. Cognitive, Noncognitive, and Environmental Correlates of Mechanical Ingenuity ... | 199 |
| PAINE, FRANK T., NASH, ALLAN N., HILLE, STANLEY J., & BRUNNER, G. ALLEN. Consumer Attitudes toward Auto versus Public Transport Alternatives | 472 |
| PASSINI, FRANK T., & NORMAN, WARREN T. Ratee Relevance in Peer Nominations | 185 |
| PENNINGTON, ALLAN L. See PETERSON, ROBERT A. | |
| PETERSON, ROBERT A., & PENNINGTON, ALLAN L. SVIB Interests and Product Preferences | 304 |
| POOCK, GARY K. Color Coding Effects in Compatible and Noncompatible Display-Control Arrangements | 301 |
| PORTUGAL, STEPHEN M. See ROTTER, GEORGE S. | |
| POULTON, E. C. Assymetrical Transfer in Reading Texts Produced by Teleprinter and by Typewriter | 224 |
| POULTON, E. C. Skimming Lists of Food Ingredients Printed in Different Brightness Contrasts | 498 |
| POULTON, E. C. Skimming Lists of Food Ingredients Printed in Different Sizes | 55 |
| RAUBENHEIMER, I. VAN W. See STARRY, ALLAN R. | |
| RHINEHART, J. B., BARRELL, R. P. DeWOLFE, A. S., GRIFFIN, J. E., & SPANER, E. E. Comparative Study of Need Satisfactions in Governmental and Business Hierarchies | 230 |
| RHODES, FEN. See MCCLELLAND, JAMES N. | |
| ROSE, J. See GOODMAN, P. | |
| ROTTER, GEORGE S., & PORTUGAL, STEPHEN M. Group and Individual Effects in Problem Solving .. | 338 |
| RUBIN, IRWIN M. See WAINER, HERBERT A. | |
| SALEH, S. D., & GRYGIER, T. G. Psychodynamics of Intrinsic and Extrinsic Job Orientation | 446 |
| SCHAEFER, CHARLES E. See ANASTASI, ANNE. | |
| SCHMIDT, FRANK, & TIFFIN, JOSEPH. Distortion of Drivers' Estimates of Automobile Speed as a Function of Speed Adaptation | 536 |
| SCHMITT, DAVID R. Punitive Supervision and Productivity: An Experimental Analog | 118 |
| SCHWAB, DONALD P., & HENEMAN, HERBERT G., III. Relationship between Interview Structure and Interinterviewer Reliability in an Employment Situation | 214 |
| SHEPPS, R. RONALD See TANOFKY, ROBERT | |
| SIMMONDS, D. C. V. See BROWN, I. D. | |
| SIMON, J. RICHARD, & SMALL, A. M., JR. Processing Auditory Information Interference from an Irrelevant Cue | 433 |
| SLOVIC, PAUL. Analyzing the Expert Judge: A Descriptive Study of a Stockbroker's Decision Processes | 255 |
| SMALL, A. M., JR. See SIMON, J. RICHARD. | |
| SPANER, F. E. See RHINEHART, J. B. | |
| STAGNER, ROSS. Corporate Decision Making: An Empirical Study | 1 |
| STARRY, ALLAN R., RAUBENHEIMER, I. VAN W., & TESSER, ABRAHAM. Stability Ratings as Classifiers of Life History Item Retest Reliability | 14 |
| STONE, THOMAS K., & ATHELSTAN, GARY T. The SVIB for Women and Demographic Variables in the Prediction of Occupational Tenure | 408 |
| STRICKER, LAWRENCE J. See Monograph Listing. | |
| STRUB, MICHAEL H. Experienced and Prior Probability in a Complex Decision Task | 112 |
| SWERTLOFF, ARNOLD B. See BOHR, RONALD H. | |
| TANOFKY, ROBERT, SHEPPS, R. RONALD, & O'NEILL, PAUL J. Pattern Analysis of Biographical Predictors of Success as an Insurance Salesman | 136 |
| TAYLOR, WILLIAM H., JR. See CARROLL, STEPHEN J., JR. | |
| TESSER, ABRAHAM. See STARRY, ALLAN R. | |
| THORNTON, CARL L. See BARRETT, GERALD V. | |
| THORNTON, GEORGE C., III. Image of Industrial Psychology among Personnel Administrators | 436 |
| THUMIN, FRED J. MMPI Scores as Related to Age, Education, and Intelligence among Male Job Applicants | 404 |
| TICKNER, A. H. See BROWN, I. D. | |

| | |
|---|-----|
| TIFFIN, JOSEPH. See SCHMIDT, FRANK. | |
| VERY, PHILIP S., & ZANNINI, JOSEPH A. Relation between Birth Order and Being a Beautician..... | 149 |
| WAINER, HERBERT A., & RUBIN, IRWIN M. Motivation of Research and Development Entrepreneurs: Determinants of Company Success..... | 178 |
| WATERS, CARRIE WHERRY. See WATERS, L. K. | |
| WATERS, L. K., & WATERS, CARRIE WHERRY. Correlates of Job Satisfaction and Job Dissatisfaction among Female Clerical Workers..... | 388 |
| WEBB, SAM C. See MARKS, EDMOND. | |
| WEST, LEONARD J. Fatigue and Performance Variability among Typists..... | 80 |
| WIENER, EARL L. See ATTWOOD, DENNIS A. | |
| WIENS, ARTHUR N., JACKSON, RUSSELL H., MANAUGH, THOMAS S., & MATARAZZO, JOSEPH D. Communication Length as an Index of Communicator Attitude: A Replication..... | 264 |
| WINICK, DARVIN. See Dubin, Jerry. | |
| WOLLOWICK, HERBERT B., & McNAMARA, W. J. Relationship of the Components of an Assessment Center to Management Success..... | 348 |
| WRIGHT, JACK M., & LEMLY, J. H. Attitudes of Professional Appraisers toward Aesthetics and Appraisal Activities..... | 531 |
| ZANNINI, JOSEPH S. See VERY, PHILIP S. | |

MONOGRAPHS

| | |
|---|--|
| BOUCHARD, THOMAS J., JR. Personality, Problem-Solving Procedure, and Performance in Small Groups. No. 1, Part 2. | |
| GRAEN, GEORGE. Instrumentality Theory of Work Motivation: Some Experimental Results and Suggested Modifications. No. 2, Part 2. | |
| STRICKER, LAWRENCE J. "Test-Wiseness" on Personality Scales. No. 3, Part 2. | |

journal of applied psychology

Table of Contents

| | | |
|--|---|--------|
| Corporate Decision Making: An Empirical Study. | Ross Stagner | 1 |
| Stability Ratings as Classifiers of Life History Item Retest Reliability. | Allan R. Starry, I. van W. Raubenheimer, and Abraham Tesser | 14 |
| Speed and Practice: Effects on Negro and White Test Performances. | Jerry A. Dubin, Hobart Osburn, and Darvin M. Winick | 19 |
| Contributions of the Interview to Assessment of Management Potential. | Donald L. Grant and Douglas W. Bray | 24 |
| Consistency and Generalizability of Intraindividual Variability. | Ralph F. Berdie | 35 |
| Comparative Reliability of Picture Form and Verbal Form Interest Inventories. | Richard Blake | 42 |
| Complex Vigilance: Relevant and Irrelevant Signals. | John L. Goldstein, William A. Johnston, and William C. Howell | 45 |
| Prediction of Job Success for Hospital Aides and Orderlies from IMPI Scores and Personal History Data. | James N. McClelland and Fen Rhodes | 49 |
| Trimming Lists of Food Ingredients Printed in Different Sizes. | E. C. Poulton | 55 |
| Knowledge of Score and Goal Level as Determinants of Work Rate. | Edwin A. Locke and Judith F. Bryan | 59 |
| 4F-Quadrant Comment. | Norman M. Abrahams | 66 |
| Moderation of a Moderator Technique. | Quinn McNemar | 69 |
| Multidimensional Analysis of Worker-Oriented and Job-Oriented Traits. | John Charles Allen | 73 |
| Fatigue and Performance Variability among Typists. | Leonard J. West | 80 |
| List of Manuscripts Accepted. | | 72, 79 |

Editor

KENNETH E. CLARK
University of Rochester

Consulting Editors

GEORGE E. BRIGGS
Ohio State University

MARVIN D. DUNNETTE
University of Minnesota

NORMAN FREDERIKSEN
Educational Testing Service

LEONARD D. GOODSTEIN
University of Cincinnati

EDWIN R. HENRY
University of Rochester

JOHN HOLLAND
American College Testing
Program

CLIFFORD E. JURGENSEN
Minneapolis Gas Company

DONALD G. MARQUIS
Massachusetts Institute of
Technology

LAURENCE S. MCCAUGHYAN
University of Houston

QUINN MCNEMAR
University of Texas

WARREN T. NORMAN
University of Michigan

HAROLD F. ROTHE
Beloit Corporation

THOMAS A. RYAN
Cornell University

ALEXANDER G. WESMAN
Psychological Corporation

CLARK L. WILSON
Instituto Centroamericano de
Administración de Empresas

This journal gives primary consideration to original investigations in any field of applied psychology except clinical psychology, although a descriptive or theoretical article may be accepted if it represents a special contribution in an applied field. Quantitative investigations of interest or value to psychologists working in the following broad fields will be considered: personnel research in business, industry, and government; engineering psychology; industrial working conditions; research on opinion and morale factors; job analysis and classification research; market and advertising research; vocational and educational prognosis, diagnosis, and guidance at the secondary and college level.

Manuscripts must be accompanied by an abstract of 100-120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or they may be obtained from the Editor or from the APA Central Office.

Manuscripts should be addressed to the Editor:

Dr. Kenneth E. Clark, Dean
College of Arts and Science
302 Morey Hall
University of Rochester
Rochester, N. Y. 14627

Authors are strongly advised to follow the general directions given in the *Publication Manual of the American Psychological Association* (1967 Revision). Special attention should be given to the section on the preparation of the references (p. 39). In reference lists, give journal titles in full; do not abbreviate. *All copy must be double spaced, including the references.* All manuscripts should be submitted in *duplicate*, one of which should be an original typed copy; author's name should appear only on title page. Dittoed and mimeographed copies are not acceptable and will not be considered. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail and to check carefully the typing of the final copy.

The following policies govern reprinting or other forms of reproduction of APA-copyrighted journal materials. Permission by the American Psychological Association must be requested for reprinting all tables, figures, and text of more than 500 words (total from one article) in length. There will be a charge of \$5.00 per page for such permission except when the requester is the author of the article to be reprinted, or the article is to be reproduced in a limited number of copies solely for instructional purposes.

Published bimonthly by the
American Psychological Association
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W.
Washington, D. C. 20036

\$10.00 per volume

\$2.00 per issue

HELEN ORR
Managing Editor

CLAIRE PILLIOD
Editorial Assistant

HENRY S. SALMON
Advertising and
Business Manager

ROBERT D. GORMAN
Subscription Manager

Subscriptions, orders, and business communications should be addressed to the American Psychological Association, 1200 Seventeenth St., N. W., Washington, D. C. 20036. Address changes must reach the subscription office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Second-class postage paid at Lancaster, Pa.

CORPORATE DECISION MAKING:

AN EMPIRICAL STUDY¹

ROSS STAGNER²

Wayne State University

A questionnaire regarding corporate decision-making (dm) practices was mailed to 500 vice-presidents of 125 large firms. Response rate was 50%. Data indicate that many goals other than profit maximization are important in decisions, and that estimates of marginal costs and profits are not always carefully made. Profitability and executive satisfaction with decision-making practices are positively correlated. Factor analysis reveals at least three important dimensions of dm process: managerial cohesiveness, formal procedures in dm, and centralization-decentralization. Factor scores derived from these factors were significantly different for firms in top and bottom thirds on profitability. However, these scores did not predict increases or decreases in profitability over a 7-yr. time span. Interpretation favors the view of the corporation as a coalition, with social role and personal bias of the executive affecting his decisions. Participative practices are supported as both satisfying and profitable.

Most of the literature on decision making in large corporations is of a highly abstract, theoretical, normative type. It sets forth, with impressive mathematical treatment, the decision processes in which corporation executives should engage if a number of quite unrealistic assumptions can be met. The articles have been characterized, perhaps unkindly, as our modern version of "how many angels can dance on the head of a pin?"

A second category, including most of the remaining publications on corporate decision making, includes memoirs of successful executives. Like case studies in the field of clinical psychology, these sometimes offer intriguing hypotheses about the process under investigation, but no acceptable data to support the proposed theory.

The field has been dominated by classical economic theory. The major assumptions about corporate decision making (dm) are as follows: (a) the firm is a unit; (b) the firm acts to maximize profit; (c) the firm is completely informed about alternative courses of action, consequences of each alternative, and the probabilities of these consequences (Simon, 1959). Some years of study of industrial conflict (Stagner, 1956) were convincing that these were contrary to fact. In an industrial dispute, top executives usually produce recommendations for corporate action, many of which are mutually contradictory, and the settlement of the dispute inside management often is nearly as difficult as that with the union. Second, the firm often acts on power considerations, or even on the basis of maintaining a public image, rather than on profit considerations. Third, members of the firm are often woefully ignorant about alternative courses of action and their probable consequences.

In recent years there has been some increase in empirical research on corporate dm, most

¹ This research was financed by a grant from the Ford Foundation for the academic year 1963-1964. The Foundation is in no way responsible for the contents of this report.

² Requests for reprints should be sent to the author, Department of Psychology, Wayne State University, Detroit, Michigan 48202.

of which has had the effect of further shaking confidence in the relevance of the classical assumptions. Simon (1959, 1960) has cited numerous instances in which the unity of the firm was a fiction. White (1961) documents many conflicts at the executive level, especially between functional departments. Dalton (1959) provides intriguing instances of the conflict between outlying branches and a central corporate office. Stagner (1965) offers various examples of disputes between vice-presidents in large- and medium-sized corporations. March (1965) shows that a firm has some characteristics of a political coalition, composed of conflicting subunits.

With respect to profit maximization as a goal, the empirical data also lead to rejection of the assumption. Soelberg (1967) has stressed the importance of individual goals which may have nothing to do with profit maximization. Simon (1960) points out that most executives accept a "satisficing" policy rather than an optimizing alternative. Stagner (1965) has shown that suboptimization may be quite common, as when a corporate policy is a compromise between what is optimal for a subunit and what is optimal for the entire firm. Feldman and Kantner (1965) point out that the alleged rule of profit maximization, if defined precisely, often fails to predict the decision made by a firm; and Mueller, Wilken, and Wood (1961) document this logical point with case studies in which an owner disregarded cost estimates in making plant location decisions.

Simon (1960) has been particularly interested in the assumption of perfect knowledge and perfect rationality in *dm*. He notes that information costs money, and most managements stop searching for alternative courses of action when they locate a "satisficing" option. Relevant case studies are those of Cyert, Simon, and Trow (1956), and Cyert, Dill, and March (1958). Marschak and Radner (1954) point out the difficulty of perfect communication from one member of the firm to another, and hence the unavailability of all the information in the *dm* process.

Psychologists have been concerned with the importance of perceptual bias in the handling of information. Cyert, Simon, and Trow (1956) gave identical case histories of a firm

to 23 executives in a training program. The selective perception of information is indicated by their answers to the question: "What is the most important problem facing the new president of this firm?" Of the sales executives, 83% named a sales problem, while only 29% of nonsales officials mentioned sales. Stagner (1965) reports instances in which production and sales managers sponsored diametrically different solutions to what was ostensibly the same problem. Zalkind and Costello (1962) have offered interpretations of the literature on perception as an aid in understanding differences in managers' choice of information to guide a decision. Appropriate to their remarks is the observation by Bowman (1961) that the "operations analyst" may fail to perceive some important item which is obvious to a working manager. Bowman advocates use of varied information sources to minimize such oversights.

Much of the "information" used in *dm* is biased by executive wishes and expectations. Cyert, Dill, and March (1958) suggest that staff personnel first decide whether the idea is good, then marshal data to support their view. Their paper gives a detailed account of the data-gathering process and cost estimates for installation of an electronic data processing unit in a medium-sized corporation. It sounds very impressive until they quote a staff member as saying, "In the final analysis, if anybody brings up an item of cost we haven't thought of, we can balance it by making another source of savings tangible [p. 340]." Similarly, Stagner (1965) quotes a corporate vice-president on the question of cost figures: "The salesmen handling this line wanted to have unit cost data. I opposed giving it to them, partly because they might unintentionally reveal it to a competitor, but more because these cost figures are in some respects artificial [p. 17]." Thus the fancy mathematical solutions developed by the operations research staff apparently incorporate highly subjective estimates of various cost factors.

Not only is the information biased; in many instances it is ignored. Consider the instance in which careful market research led to "a proposal to set up three installations, each costing \$5 million, two in the United States and one in western Europe. The head of the

English subsidiary and the head of the French subsidiary got into a feud over which would get the European unit. After considerable negotiating, the American controlling executives decided to put one each in France and England [Stagner, 1965, p. 17].” In this instance detailed staff work had indicated that one installation was adequate to the foreseeable European market. The key issue was the relative status and power of the two European executives, and \$5 million was the “side payment” to keep peace in the organization. This is an instance of “suboptimization,” a compromise between optimum for the subunit and optimum for the entire firm. It suggests that profit is only one of many goals which determine corporate decisions.

Economic theorists have not ignored absolutely the conflict of their assumptions with empirical reality. In an intriguing effort to incorporate some of these observations, Harsanyi (1962) suggests that optimizing equations be rewritten to include such factors as the opportunity cost to A of getting and using power, the cost to B of refusing to yield to A, and the personal affection of B for A. Obviously such a formulation deviates rather far from the simple profit-maximization approach.

These observations suggest that there is an urgent need for research on corporate dm processes which is theoretically based but reasonably close to empirical reality. This study offers a beginning on that task.

METHOD

A Theoretical Position

It seems appropriate to approach the problem of corporate dm by making the following assumptions: (a) corporate policy is established by persons occupying certain role positions in the organization; (b) the behavior of these persons is determined in part by role prescriptions, and in part by personal motives; (c) perceptions of corporate resources, alternative actions, utilities, and probabilities of outcomes will be affected by role-induced experiences and by personal experiences; (d) policy proposals by different executives will reflect these divergent perceptions and motivations; and (e) corporate decisions will represent compromises among these, affected by the power of proponents as well as by logic and realistic data.

Research on this kind of complex phenomenon really should involve analysis of all relevant documents, video recordings of all relevant conversations and conferences, depth explorations of each execu-

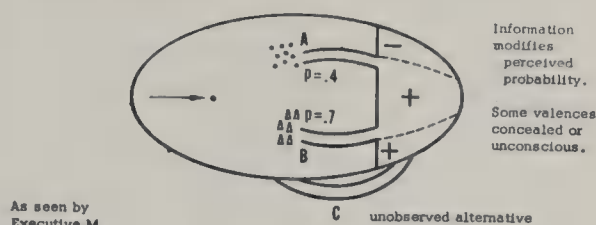


FIG. 1. A corporate decision situation as perceived by one executive.

tive to ascertain conscious and unconscious desires, symbols, etc., affecting his policy preferences, and an assessment of the power fields of the various executives. For many reasons such research is not presently feasible. As a substitute we may rely on reports of dm from participant executives, and try to ferret out the process from these subjective data. The task is difficult. Phenomenological reports are slippery even when made by trained observers having no aspirations likely to bias their observations. Industrial executives are, for this purpose, untrained observers, and it is assumed that they will introduce biases into their reports. Nevertheless, this seems the only suitable source of data at this time for a study of high-level dm processes.

Our approach proposes that an individual executive view a problem situation as sketched in Fig. 1. He is fully aware of a major corporate goal (which need not be profit; it may be competitive standing, public “image,” political advantage, etc.) and he is aware that attainment of this goal is impeded by some difficulty; there is a discrepancy between the existing state and the preferred state of affairs. He is also affected by, but not necessarily conscious of, various other positive and negative goals. Some of these relate to his department, division, or staff position in the corporation; others to his personal power, prestige, or profit. He sees possible courses of action, positive and negative utilities, and probabilities of various outcomes.

Another executive perceives the situation in a slightly different way (Fig. 2); his alternatives, utilities, and probabilities reflect his role position and his personal needs. Discussion of the issue by the two persons results in overlapping phenomenal fields (Fig. 3), but the theory assumes that these never coincide perfectly.

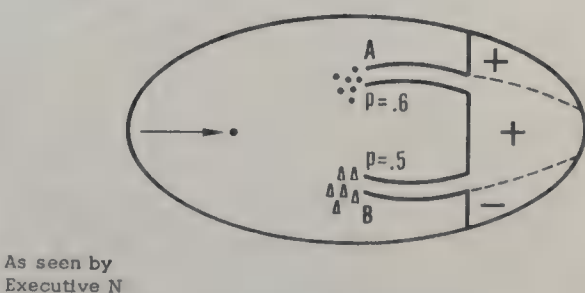


FIG. 2. The decision situation of Fig. 1 as perceived by another executive.

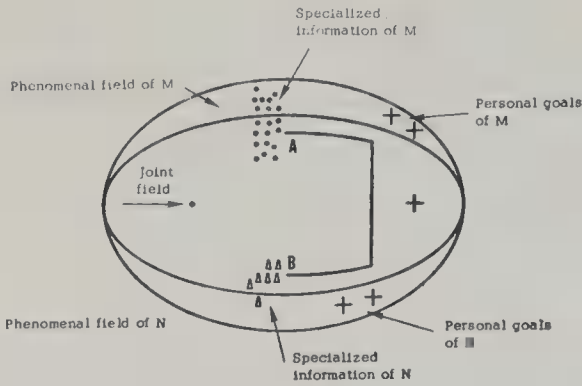


FIG. 3. The joint field of the two executives.

Adjustment of the remaining differences may be brought about by (a) cognitive modification (data which induce changes in one phenomenal field) or (b) dynamic modification, involving "side payments" or rewards and punishments which bring about modification of valences. The solution to the problem may result from the introduction of a power field (e.g., the chief executive, Fig. 4) which removes all alternatives but one. It may, on the other hand, be a compromise between the solutions preferred by the subordinates, as in Figure 5. In extreme cases the solution involves replacement of one or more executives.

The dm in such a theoretical context will be affected by various objective environmental constraints (the "foreign hull" of Lewin's lifespace), by subjective constraints (the executive's refusal to consider some alternatives), by the internal pattern of communications and authority relations, by the traditions of the firm, etc. Each of these can be inferred at some unspecified level of precision from reports given by executives of their participation in the dm process.

Hypotheses

While this was designed as an exploratory study, the following hypotheses were set up for testing:

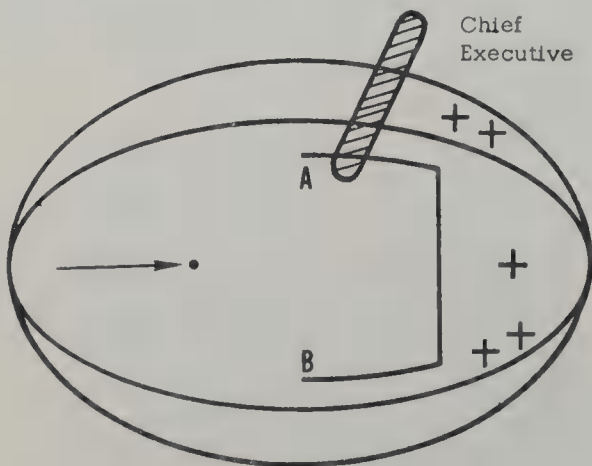


FIG. 4. Decision by blocking one alternative.

1. Profit maximization will be the only goal reported by executives.
2. All corporations make decisions in the same manner, i.e., there are no significant differences in style of dm.
3. An executive's power status within the firm does not affect his part in the dm process.
4. Profitability is unrelated to variations of manner of making decisions.
5. Personality variables are nonsignificant in the corporate dm process.

Procedure

This study develops from an earlier project (Stagner, 1965) in which unstructured interviews explored executive perceptions of the corporate dm process. Analysis of the interview material led to the formulation of 28 questions which seemed to merit quantitative analysis. These were classified into six groups, although these categories are useful primarily for noting relations to theory, not for statistical analysis. The groups, and illustrative items, were: *goal variables*: cost and profit estimates, company tradition, corporate image. *Means variables*: channels of communication, lines of authority, speed of decisions, formal routines, use of ad hoc committees, use of outside consultants. *Leadership variables*: chief executive talks to one vice-president at a time; chief executive is concerned that all be satisfied. *Role variables*: conflict between central office and divisions; vice-presidents exaggerate importance of their division. *Interaction variables*: groups among top executives; tension among top executives; discussion among all persons affected; loser (on a decision) feels defeated. *Outcome variables*: satisfaction with decision-making procedure; morale of top executives. (In addition, profits as percentage of capital and profits as percentage of sales were used as outcome measures but these were not in the questionnaire.)

The questions were formulated so that they could be answered by making check marks on a 7-step scale, to avoid the objection to forced yes-no answers. The wording of the items and the definitions of the ends of each 7-step scale, with the answers from 217 vice-presidents, are given in Table 1.

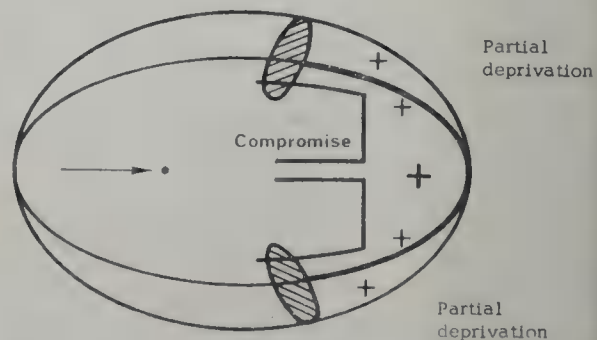


FIG. 5. Decision by compromise (partial deprivation for each executive).

TABLE 1
QUESTIONNAIRE ITEMS AND RESPONSES

| Item | Response | "Very important" |
|--|---|------------------|
| 1. Relative speed with which top-level decisions are made | : <u>2</u> : <u>4</u> : <u>22</u> : <u>48</u> : <u>39</u> : <u>73</u> : <u>29</u> : somewhat slower than similar companies considerably faster than similar companies | <u>87</u> |
| 2. Concern over formal steps in decision-making at top level (regular meetings, written records, etc.) | : <u>11</u> : <u>22</u> : <u>34</u> : <u>35</u> : <u>37</u> : <u>54</u> : <u>24</u> : much attention to formal routines little attention to form | <u>15</u> |
| 3. Estimates of cost and anticipated profit to result from a decision | : <u>62</u> : <u>47</u> : <u>45</u> : <u>32</u> : <u>18</u> : <u>11</u> : <u>2</u> : always carefully computed rough estimates only | <u>86</u> |
| 4. Discussion among all top executives | : <u>40</u> : <u>52</u> : <u>37</u> : <u>27</u> : <u>25</u> : <u>23</u> : <u>13</u> : discussions include all executives affected most include only two men at a time | <u>59</u> |
| 5. Use of a top-level policy committee or operating committee | : <u>49</u> : <u>15</u> : <u>13</u> : <u>19</u> : <u>18</u> : <u>58</u> : <u>45</u> : we have none it merely approves decisions already made it is an active decision-making apparatus | <u>36</u> |
| 6. Use of ad hoc or special committees for single projects | : <u>29</u> : <u>27</u> : <u>41</u> : <u>37</u> : <u>25</u> : <u>38</u> : <u>20</u> : common practice usually one person, not a committee | <u>7</u> |
| 7. Tendency of each vice-president to exaggerate importance of his own area | : <u>2</u> : <u>4</u> : <u>4</u> : <u>25</u> : <u>21</u> : <u>98</u> : <u>63</u> : serious problem here not a serious problem | <u>9</u> |
| 8. Social interaction of top executives outside office hours (for nonbusiness purposes) | : <u>15</u> : <u>16</u> : <u>27</u> : <u>44</u> : <u>25</u> : <u>70</u> : <u>20</u> : frequent rare | <u>1</u> |
| 9. Concern of chief executive for detailed information on which to base decision | : <u>37</u> : <u>41</u> : <u>30</u> : <u>22</u> : <u>23</u> : <u>47</u> : <u>17</u> : wants substantial detail prefers broad outlines | <u>40</u> |
| 10. Concern of chief executive that all executives are satisfied with the decision | : <u>23</u> : <u>35</u> : <u>44</u> : <u>37</u> : <u>31</u> : <u>38</u> : <u>9</u> : considerable concern minor concern | <u>18</u> |
| 11. Concern with "going through channels" | : <u>7</u> : <u>22</u> : <u>32</u> : <u>39</u> : <u>47</u> : <u>52</u> : <u>18</u> : relatively little attention to this communications always observe channels | <u>24</u> |
| 12. Importance attached to company tradition and past policies | : <u>19</u> : <u>36</u> : <u>43</u> : <u>36</u> : <u>35</u> : <u>33</u> : <u>15</u> : not much; easy to break tradition considerable weight attached to tradition | <u>11</u> |
| 13. Concern of chief executive for "unanimous agreement" | : <u>1</u> : <u>7</u> : <u>8</u> : <u>19</u> : <u>31</u> : <u>101</u> : <u>50</u> : very reluctant to confirm policy if anyone opposes approves policy when he sees a clear majority | <u>17</u> |
| 14. Reaction of executives to decisions which go against their preference | : <u>2</u> : <u>4</u> : <u>4</u> : <u>26</u> : <u>37</u> : <u>109</u> : <u>35</u> : often feel "defeated" in such a case usually accept decision without feeling "defeat" | <u>14</u> |

TABLE 1—(Continued)

| Item | Response | "Very important" |
|--|---|------------------|
| 15. Preferred style of chief executive | : 20 : 24 : 22 : 36 : 25 : 66 : 24 : talk with one man at a time talk with all interested men together | 24 |
| 16. Normal operation of divisions within company | : 43 : 53 : 44 : 31 : 22 : 15 : 9 : most divisions highly independent little divisional autonomy | 38 |
| 17. Tendency of chief executive to give the "losing" executive some other concession to make him feel better | : 3 : 10 : 18 : 34 : 44 : 84 : 24 : often happens never happens | 1 |
| 18. Chief executive's preference for division heads to be partisan of division or look at company as a whole | : 5 : 10 : 19 : 30 : 28 : 82 : 43 : prefers strong division partisanship wants all to think chiefly of company, not division | 37 |
| 19. Use of outside consultants | : 5 : 19 : 31 : 31 : 42 : 54 : 35 : used on most important issues rarely used by this company | 4 |
| 20. Clear lines of authority | : 31 : 63 : 47 : 26 : 21 : 21 : 8 : everyone knows and respects lines of authority lines of authority ambiguous, often ignored | 84 |
| 21. Importance attached to company "image" as seen by public | : 10 : 36 : 33 : 61 : 29 : 40 : 8 : often outweighs cost factors would have little effect | 27 |
| 22. Importance of personalities in decisions at this level | : 6 : 18 : 39 : 35 : 38 : 76 : 17 : vigorous, persuasive individual often wins point importance of function to company usually decisive | 21 |
| 23. Importance of divisional vs. central office disagreements | : 27 : 61 : 47 : 47 : 13 : 17 : 5 : few such disagreements these are fairly common | 10 |
| 24. Ability of strong divisions to get their own way | : 7 : 40 : 52 : 31 : 30 : 42 : 15 : strong divisions win if deeply concerned no differences among divisions | 9 |
| 25. Groups within top executive echelon | : 2 : 13 : 14 : 14 : 35 : 89 : 50 : some men habitually vote together as a group no tendency toward alignments | 15 |
| 26. Amount of tension within top executive group over a difficult decision | : 22 : 59 : 47 : 36 : 34 : 14 : 5 : no tension or personal feelings tension sometimes high, personal frictions | 19 |
| 27. Your estimate of <i>morale</i> of top executive echelon | : 4 : 4 : 10 : 17 : 32 : 105 : 45 : low high | 81 |
| 28. Your feeling of satisfaction with the way these decisions are handled | : 48 : 70 : 42 : 18 : 23 : 11 : 5 : very well satisfied some satisfaction | 12 |

Sample of Firms

The questionnaire as described was sent to 500 persons at the vice-presidential level (VPs). Sampling was as follows: from the *Fortune Magazine* list of 500 largest American corporations (July, 1963), 125 were selected by taking every fourth name.³ From Standard and Poor's Registry of Directors were obtained the names of four VPs (including when necessary treasurers, controllers, etc., if not enough VPs were listed). Thus, 500 questionnaires were mailed. These were coded to identify firms but not individual respondents.

Returns were received from about 260 individuals. All four officials replied from 6 firms; three from 29 firms; two from 46 firms; and only one from each of 28 firms; but a number of these were incomplete. The final analysis is based on 217 responses from 109 firms.

RESULTS

The distribution of 217 executives from 109 firms, in terms of their answers to specific items, is given in Table 1. Some interest will attach to these in the light of points made above. For example, it is obvious that in many firms, cost and marginal profit estimates are not carefully made (Item 3). In fact, a substantial number (28%) indicated that "rough estimates" were made of such variables. Table 1 also indicates that company "image" may outweigh profit considerations (Item 21) and adherence to tradition may also be an important value (Item 12). These data merely confirm observations already made regarding the importance of goals other than profit maximization. They lead to rejection of Hypothesis 1 (that profit will be the only goal reported).

The first seven columns in Table 1 show the number of VPs checking at each of the seven steps on the answer scale. Column 8 gives the number checking that item as very important. Various computations involving this datum led to absolutely no meaningful results, and the conclusion follows that these respondents were not good judges of the relative importance of various items in relation to dm procedure or outcome.

³ A few substitutions were made, e.g., when the above process turned up atypical organizations such as a large farmers' cooperative. In such instances the replacement was the next firm on the list.

TABLE 2

CORRELATIONS BETWEEN PAIRS OF EXECUTIVES
DESCRIBING THE DECISION-MAKING PROCESS

| Group | Mean z' | Mean r |
|------------------------------|-----------|----------|
| A (intra-firm) | .463* | .433* |
| B (random multiple-response) | .330 | .319 |
| C (random single-response) | .310 | .300 |

Note.— t (A-B), 2.78, $p < .05$; t (A-C), 4.95, $p < .05$; t (B-C), 0.59, ns .

Reliability of Data

To talk about the dm process in a firm, it must be shown that it can be described with sufficient precision as to be different from some other firm. To test this, the authors took 52 pairs of men from the multiple-responding firms and computed profile correlations (the pattern of responses on one questionnaire against that on another). Only one pair was taken from any given firm (Group A, Table 2). Then each of these 52 was paired against a man from a different firm. These are shown as Group B. As another control we took 190 random pairs from firms sending in only a single response (Group C). For pairs within the same firm the actual range of correlations was from +.79 (rather high agreement) to +.01 (no agreement at all). The mean (Table 2) is +.43. For Group B (same men, paired across firms) the range is +.73 to −.27, with a mean of +.319. In Group C (random pairs from firms with one respondent) the range is +.86 to −.53; mean, +.30. The A-B and A-C differences meet the .05 level for a two-tailed test. Thus we can say conservatively that two VPs from the same firm will agree more in their responses than two from different firms, and the hypothesis of an observable event (the dm process) is sustained. (Conversely, it is a bit discouraging that the agreement within the firm is not higher; it suggests that the unity of the firm is even less than suggested earlier. We thus reject Hypothesis 2 (that the dm process is the same in all firms.)

An examination of these profile correlations suggests that in some firms there is high communication and mutual understanding, while in others there is fractionation and dissociation. For one group of three VPs in the same

TABLE 3
ITEMS RECEIVING HIGHEST LOADINGS ON THE
UNROTATED MATRIX, FACTOR I

| Item | Loading |
|---|---------|
| 28. Your feeling of satisfaction with way decisions are handled (very well) | .77 |
| 27. Your estimate of top executive morale (high) | .73 |
| 20. Clear lines of authority (yes) | .65 |
| 4. Discussion among all top executives (all incl.) | .59 |
| 26. Tension within top executive group (rare) | .57 |
| 25. Groups within top executive echelon (rare) | .54 |
| 11. Concern with going through channels (yes) | .54 |

Note.—The answer given in parentheses defines the positive end of the 7-step scale. See Table 1 for the complete items. For *N* = 217, a correlation of .30 is significant at the .01 level.

firm, the correlations are .66, .79, and .64; for another set of three, the figures are .18, .13, and .38. This fact fits with other data to be mentioned later.

Dimensions of the DM Process

One major concern of the study was to identify styles of dm in corporations, and the tactic utilized for this purpose was factor analysis. Each of the 28 questionnaire items, plus size, profit on sales, and profit on capital, was correlated with all the others, and subjected to a principal axes analysis. Seven factors emerged, of which four had variances above 1.0. Only the first two seemed to make sense in terms of everyday knowledge of corporation functioning. Factor I is heavily loaded on executive morale (Table 3). This seems plausible in view of the number of

TABLE 4
ITEMS RECEIVING HIGHEST LOADINGS ON THE
UNROTATED MATRIX, FACTOR II

| Item | Loading |
|--|---------|
| 30. Profit as % of sales | .47 |
| 31. Profit as % of capital | .42 |
| 2. Formal routines (yes) | .39 |
| 6. Ad hoc committees (common) | .34 |
| 19. Outside consultants (often) | .34 |
| 1. Speed of decisions (slow) | .33 |
| 7. Losing executive feels defeated (yes) | .31 |

questions asked which would bear, in one way or another, on the satisfaction of executives with the dm process. Teachers of industrial psychology and management courses will be pleased to note that “going through channels” and “clear lines of authority” favor high morale. Profitability is positively but not heavily loaded on this factor.

Factor II has its highest loadings on the two profitability indexes (Table 4). In addition, it includes several “means” items: “concern over formal steps,” “use of ad hoc committees,” “outside consultants,” and relatively slow decisions. One possible interpretation, based partly on independent knowledge of the firms, is that this factor weights profitability based on strong decentralized divisions, as opposed to a cohesive central administration. Further support for this view derives from Table 5, showing items which reverse sign from Factor I to II. It will be noted that the loadings in Factor I point to an inte-

TABLE 5
ITEMS WHICH REVERSE SIGN FROM
FACTOR I TO FACTOR II

| Item | Answer positively loaded on | |
|---|-----------------------------|---------------------|
| | I | II |
| Each VP exaggerates importance of his division | No | Yes |
| Executives feel “defeated” if losing decision | No | Yes |
| Chief executive gives some other concession to loser | No | Yes |
| Lines of authority clear | Yes | Maybe |
| Corporate function more important than vigorous personality | Yes | Not always |
| Conflicts of divisions vs. central office | Few | Common |
| Strong divisions get own way | Not usual | If deeply concerned |
| Groups within top echelon of executives | No | Some |
| Tension among executives over tough decisions | No | Sometimes high |

TABLE 6

ITEMS RECEIVING HIGHEST LOADINGS ON THE
ROTATED MATRIX, FACTOR I'

| Item | Loading |
|---|---------|
| 27. Estimate of top executive morale (high) | .67 |
| 26. Tension at top (low) | .63 |
| 14. Losing executive feels defeated (no) | .58 |
| 28. Your satisfaction with procedure (high) | .57 |
| 23. Conflicts between central office and divisions (rare) | .56 |
| 7. Vice-presidents exaggerate area importance (not serious) | .53 |
| 25. Groups at top (no) | .44 |

grated structure, while II points to a number of autonomous units under a single corporate roof. Both factors point to rejection of Hypothesis 4 (profitability not related to manner of dm).

Most factor analysts would argue that rotation of factors to simple structure provides the best approach to identifying meaningful dimensions in a mass of correlational data. The seven factors were rotated by Varimax, giving quite a different pattern from the unrotated structure.⁴

Factor I' (Table 6) confirms the suggestion of a dimension of managerial cohesiveness. Firms high on this factor have managers skilled at working together; but the items give us no hint of how this coordination was achieved.

Factor II' (Table 7) might be called "formality in dm"; it has elements of participative management style and also essential bureaucratic procedures. The style of the chief executive seems to be important here in keeping a "tight ship" but at the same time listening to all concerned and maintaining high morale. It is a little surprising that, statistically, this is independent of Factor I'.

Factor III' has precisely two items loading significantly—profit on sales and profit on capital. This seems to confirm the factor analyst's belief that his procedure can extract a logically independent factor even if, in the raw

⁴ The author wishes to acknowledge the generous assistance provided by the staff of the Wayne State University Computing Center in adapting programs for this purpose, and the assistance of D. R. Jacobs, in cross-checking many details in the data analysis.

TABLE 7

ITEMS RECEIVING HIGHEST LOADINGS ON THE
ROTATED MATRIX, FACTOR II'

| Item | Loading |
|--|---------|
| 4. Discussion among all executives affected (yes) | .62 |
| 10. Chief shows concern that all are satisfied (yes) | .52 |
| 15. Chief talks with one executive at a time (no) | .52 |
| 11. Go through channels (yes) | .52 |
| 20. Clear lines of authority (Yes) | .51 |
| 28. Your satisfaction with procedure (high) | .51 |
| 2. Formal routines in decision-making (yes) | .47 |

data, it is thoroughly mixed in with other items.

Factor IV' seems to be a "fragmentation" or decentralization dimension (Table 8). It is plausible that such a dimension would exist in a population of corporations, but puzzling that it is independent of I'.

Factor V' seems to represent a group of firms with highly personalized management, by which is meant that personalities may weigh more heavily than organization. However this factor accounts for only 10% of the common variance, which fits with other reports that corporate structure and power, not personality as such, determine dm outcomes. The two remaining factors were discarded because they had few significant loadings.

DM and Corporate Outcomes

Major goals of corporation executives, as postulated, include profits and competitive

TABLE 8

ITEMS RECEIVING HIGHEST LOADINGS ON
ROTATED MATRIX, FACTOR IV'

| Item | Loading |
|---|---------|
| 18. Chief executive prefers division heads to be partisan of division (Yes) | .48 |
| 16. Normal operation of divisions (highly independent) | .46 |
| 24. Ability of strong divisions to get own way (win if deeply concerned) | .45 |
| 12. Importance attached to company tradition and past policies (not much) | .36 |
| 22. Importance of personalities in decisions (vigorous person often wins) | .32 |

TABLE 9
PROFIT AS PERCENT ON SALES

| | |
|-----|---|
| 2. | Concern over formal steps in decision making at top level (regular meetings, written records, etc.) (much attention to formal routines) |
| 3. | Estimates of cost and anticipated profit to results from a decision (always carefully computed) |
| 4. | Discussion among all top executives (yes) |
| 8. | Social interaction of top executives outside office hours (for nonbusiness purposes) (frequent) |
| 11. | Concern with "going through channels" (always observe) |
| 20. | Clear lines of authority (everyone knows and respects) |
| 21. | Importance attached to company "image" as seen by public (often outweighs cost factors) |
| 28. | Your feeling of satisfaction with the way these decisions are handled (very well satisfied) |

Note.—Items differentiating at .01 level.

stature. The variables of profit as percentage of sales, profit as percentage of capital, and size, represent indexes of such goal-achievement. Does the type of dm activity within a firm have any relevance for such indexes of success?

We may first look at some specific items from the questionnaire, and then at factor scores based on the dimensional analysis. An item analysis was carried out by taking the top third and bottom third of all firms on each of the three outcome variables, and running t-tests on the difference in mean response to each item.

Profit on sales. The eight items which differentiated at the .01 level between firms most and least profitable are shown in Table 9. Some of these would be expected (costs and marginal profits are carefully estimated, as Adam Smith would have urged). Bureaucratic routines are also high. However, there is evidence of the importance of interaction variables (discussions, outside socializing) which suggests that formal structure by itself is not enough. Some observers will be amused by the fact that "company image may outweigh cost factors" leads to more profitability,

not to losses. And in this context one is not sure whether executive satisfaction helps earn profits, or whether firms that earn profits have satisfied VPs.

Profit on capital. Six items meet the .01 criterion for profit on capital. Only minor differences are reflected if we compare these (Table 10) with those in Table 9. Profit on capital may be a bit more closely related to bureaucratic routines. However, much of this apparent difference would disappear if we published the items differentiating at the .05 level; in general, conditions favoring profit on sales are also those which favor profit as a percentage of capital. It should be noted that ranking for profit on sales correlates .71 (in this sample) with ranking for profit on capital.

Size. Only five items distinguish the top and bottom thirds of the size distribution (Table 11). Most of these are plausible in the sense that we would expect, in a larger firm, that cost estimates would be more carefully made, that the chief would not ask for much detail, and so on.

The data were also analyzed to see if size functioned as a moderator variable to affect profitability/dm relationships. The exploration did not confirm the tentative hypothesis that profitable management of a smaller enterprise would follow a different pattern from that in a larger firm. In caution, it should be noted that this sample of firms was limited to

TABLE 10
PROFIT AS PERCENT ON CAPITAL

| | |
|-----|--|
| 5. | Use of a top-level policy Committee or operating committee (Yes) |
| 8. | Social interaction of top executives outside office hours (for nonbusiness purposes) (frequent) |
| 9. | Concern of chief executive for detailed information on which to base decision (wants substantial detail) |
| 15. | Preferred style of chief executive (talk with all interested men together) |
| 20. | Clear lines of authority (yes) |
| 28. | Your feeling of satisfaction with the way these decisions are handled (very well satisfied) |

Note.—Items differentiating at .01 level.

TABLE 11

ITEMS DIFFERENTIATING SMALLER FROM
LARGER FIRMS

| |
|--|
| 3. Estimates of cost and anticipated profit (carefully made) |
| 28. Your feeling of satisfaction (not very high) |
| 5. Use of top-level policy committee (yes, active) |
| 9. Concern of chief with detailed information (no) |
| 20. Clear lines of authority (yes) |
| 27. Estimated morale at top (high) |

Note.—Answer for larger firms; items significant at .05 level.

rather large corporations—and highly profitable ones, too—so that such differences may not have been identifiable within the sample studied.

Factor scores and profits. Table 12 shows the factor scores on rotated Factors I', II', IV', and V' for the top and bottom thirds on profit on sales.⁵ Similarly, Table 13 shows the mean scores for profit on capital.

Decidedly surprising is the fact that all four factors are significant in Table 13 and two in Table 12. The two nonsignificant factors in Table 12 show differences in the same direction as those in 13. It will be recalled that profitability had very small loadings on all four factors after rotation. In both Tables 12 and 13, profitability is associated with greater cohesiveness, more formality (bureaucratic routines), centralization, and tendency away from personalized management. It thus seems fair to conclude that outcomes are affected by style in dm even when some effort has been made to exclude the effects of profitability statistically.

Changes in profitability. The last, and most severe, test of any measure of real-life variables is its ability to predict outcomes at a later date. In the present instance, data were collected in the winter and spring of 1964. Since *Fortune Magazine* obligingly publishes

⁵ The factor scores used in this analysis are approximations. The heavily loaded items (shown in earlier tables) were scored for a given factor, but were not differentially weighted. The increase in precision which would have resulted would have been minimal, and the absence of a cross-validation sample argued against any need for added precision.

TABLE 12

FACTOR SCORES RELATED TO PROFITABILITY
ON SALES 1963

| Factor | High profit rank | Low profit rank | p |
|----------------------------------|------------------|-----------------|------|
| I' organizational cohesiveness | 27.52 | 31.14 | .025 |
| II' formality in decision making | 37.32 | 46.14 | .001 |
| IV' decentralization | 26.15 | 25.70 | ns |
| V' personalized management | 25.49 | 24.74 | ns |

its list of the 500 largest corporations annually, and provides rankings on profitability on sales and capital each year, we were tempted to see if our dm indexes predicted change in profitability.

Profit ranks were taken for 1958, 1963, and 1965. Change scores were computed for 1958–1963, 1958–1965, and 1963–1965. If dm indexes recorded relative managerial efficiency, perhaps they would correlate with these change scores. We therefore correlated the four factor scores with these six change scores, a total of 24 correlations. Not a single one reached the .05 level of significance. We must conclude that the hypothesis that dm predicts increase or decrease in profitability has been disconfirmed.

DISCUSSION

Data confirm the author's expectancies—and hence, perhaps, are suspect—in that they contradict the assumptions of classical theory. Individual executives and variant forms of corporate organization do have significant effects on the dm process. Profit is a major goal

TABLE 13

FACTOR SCORES RELATED TO PROFITABILITY ON
CAPITAL 1963

| Factor | High profit rank | Low profit rank | p |
|----------------------------------|------------------|-----------------|------|
| I' organizational cohesiveness | 27.36 | 30.62 | .025 |
| II' formality in decision making | 38.84 | 44.47 | .01 |
| IV' decentralization | 26.58 | 25.22 | .01 |
| V' personalized management | 26.49 | 24.27 | .001 |

but by no means the only one, and most executives agree that in some instances they reject profit in favor of some other value.

The data lend considerable support to the view of the firm as a coalition (March, 1962). Strong divisions within the company may get their way without regard to the welfare of the whole (Item 24), VPs exaggerate importance of their divisions (Item 7), some chief executives actually prefer this (Item 18), and factionalism at the top is by no means rare (Item 25). Factor IV' points to a common pattern or type of corporation which is relatively decentralized, a set of almost autonomous units under a central umbrella which may be little more than a financial holding company.

On the other hand, support is available here for the positions taken by Likert (1967) and McGregor (1960) in favor of participative management. Involvement of all executives (Item 4) and concern by the chief that all be satisfied (Item 10) are associated with high executive morale, satisfaction with the dm process, and profitability. The small, decision-making echelon at the top of a large corporation has some attributes in common with the small groups studied by experimental social psychologists, and with the committees studied by Collins and Guetzkow (1964). Communication patterns are important; and the efficiency of centralized structure, with concomitant loss of satisfaction for those on the periphery, seems to be involved in these results.

Leadership is also important. The distinction between "mediator" and "arbitrator" styles (Stagner, 1965) and the contrast in the present data between bilateral talks and wider executive participation may be related to the now traditional distinction between "consideration" and "structuring." A chief executive who wants all affected executives to participate and be satisfied is certainly closer to the "consideration" pole than are those who disregard such matters.

These data do not support the hypothesis that a vigorous personality may win a decision against opposition with a stronger power base. The social role, or control of organizational power, seems generally more significant. This conclusion, of course, requires

qualification. Most of the men contacted in the earlier study (Stagner, 1965) were judged to have vigorous, aggressive personalities. It can reasonably be assumed that respondents in this survey were similar. One does not become a VP in a large American corporation by passive-dependent behavior. Personality differences on this dimension therefore may have been relatively small, thus obscuring any significant trends.

It is important to say a word about the problem of corporate goals. We must replace the concept of a single utility, profit maximization, with the concept of multiple utilities. Industrial psychologists should recognize this as an extension of the problems encountered with personnel test validation. Reliance on a single criterion measure, such as quantity of production, or incentive earnings, proved hopelessly oversimplified. According to research findings, personnel decisions should be based on multiple criteria with minimum cutting scores on a number of predictor variables rather than a single regression equation. The personnel manager wants to hire workers who will have good absentee and tardiness records, will accept supervision, and will produce at a high level in quality and quantity. There is a point on each of these beyond which the employee is unacceptable regardless of how good he is on other aspects of performance. Thus the manager accepts a "satisficing" solution with respect to several variables.

The problem of corporate goals must be handled in the same manner. Executive decisions must balance costs against customer goodwill, efficient production against union resistance, pricing policy against political repercussions. The executive in such cases seeks a satisficing level for the other outcome measures and then attempts to maximize profit. Various executives may be more interested in maximizing for their own division than for the firm. Inevitably, optimization becomes a delusion if taken literally. Corporate dm is guided by numerous values, only one of which is profit maximization.

REFERENCES

- BOWMAN, E. H. Management decision making: Some research. *Industrial Management Review*, 1961, 3 (1), 56-63.

- COLLINS, BARRY E., & GUETZKOW, HAROLD. *A social psychology of group processes for decision-making*. New York: Wiley, 1964.
- CYERT, R. M., DILL, W. R., & MARCH, J. G. The role of expectations in business decision-making. *Administrative Science Quarterly*, 1958, 3, 307-340.
- CYERT, R. M., SIMON, H. A., & TROW, D. B. Observation of a business decision. *Journal of Business*, 1956, 29, 237-248.
- DALTON, MELVILLE. *Men who manage*. New York: Wiley, 1959.
- FELDMAN, JULIAN, & KANTER, HERSCHEL E. Organizational decision making. In J. G. March (Ed.), *Handbook of organizations*. Chicago: Rand McNally, 1965.
- FISK, GEORGE (Ed.) *The psychology of management decision*. Lund, Sweden: CWK Gleerup Publishers, 1967.
- FORTUNE. *The Fortune Directory: The 500 largest U. S. industrial corporations*. 1963, 68(1), 177-196.
- HARSANYI, JOHN C. Measurement of social power, opportunity costs, and the theory of two-person bargaining games. *Behavioral Science*, 1962, 7, 67-80.
- LIKERT, R. *The human organization: Its management and value*. New York: McGraw-Hill, 1967.
- MARCH, J. G. The business firm as a political coalition. *Journal of Politics*, 1962, 24, 662-678.
- MARCH, J. G. *Handbook of organizations*. Chicago: Rand McNally, 1965.
- MARSCHAK, J., & RADNER, R. The firm as a team. *Econometrica*, 1954, 22, 523. (Abstract).
- MCGREGOR, D. *The human side of enterprise*. New York: McGraw-Hill, 1960.
- MUELLER, E., WILKEN, A., & WOOD, M. *Location decisions and industrial mobility in Michigan*. Ann Arbor, Mich.: Institute for Social Research, University of Michigan, 1961.
- Poor's Register of Corporations, Directors, and Executives*. New York: Standard & Poor Corporation, 1963.
- SIMON, H. A. Theories of decision-making in economics and behavioral science. *American Economic Review*, 1959, 49, 253-283.
- SIMON, H. A. *The new science of management decision*. New York: Harper & Row, 1960.
- SOELBERG, P. Structure of individual goals: Implications for organizations theory. In G. Fisk (Ed.), *The psychology of management decision*. Lund, Sweden: CWK Gleerup Publishers, 1967.
- STAGNER, R. *The psychology of industrial conflict*. New York: Wiley, 1956.
- STAGNER, R. Resolving top-level managerial disagreements. *Business Topics*, 1965 (Winter), 13, 15-22.
- WHITE, H. C. Management conflict and sociometric structure. *American Journal of Sociology*, 1961, 67, 185-199.
- ZALKIND, S. S., & COSTELLO, T. W. Perception: Some recent research and implications for administration. *Administrative Science Quarterly*, 1962, 7, 218-235.

(Received September 11, 1968)

STABILITY RATINGS AS CLASSIFIERS OF LIFE HISTORY ITEM RETEST RELIABILITY

ALLAN R. STARRY¹ AND I. VAN W. RAUBENHEIMER

Measurement and Research Center, Purdue University

AND ABRAHAM TESSER

University of Georgia

Investigations concerned the prediction of fakeability and stability of validated life history items. Two studies were conducted: (a) 26 judges rated 60 (Questionnaire A) and 52 (Questionnaire B) discretely scored items for probable fakeability and stability, indexes of which were based on retest responses of 321 undergraduates over 3 mo. Validity coefficients for the fakeability and stability ratings were .53 (Questionnaire A) and .50 (Questionnaire B), and .54 (Questionnaire A) and .51 (Questionnaire B), respectively ($p < .01$). (b) 25 judges rated 88 continuously scored items for probable response stability. Correlation with actual retest stability for 106 freshmen was .47 ($p < .01$). Little difference was found in predicting stability for discrete and continuous items. The Probable Response Stability scale as a potential classifier variable of life history items as well as potential difficulties and suggestions for future research are discussed.

Current widespread use of life history items as predictors of academic and vocational criteria dictates that more information be obtained on the stability of responses to biographical questionnaires. Typically such forms are item analyzed and validated on a concurrent basis with samples of employees, students, etc., on whom criterion data are available. The assumption that responses to personal history items will remain stable over time is thus implicit when the scored questionnaire is subsequently utilized as a selection tool. Several investigators have demonstrated, however, that relatively imperfect stability is a more reasonable expectation. Their studies indicate that item characteristics (Owens, Glennon, & Albright, 1962), criterion transparency (Klein & Owens, 1965), and objectionability (Larsen, Swarthout, & Wickert, 1967) are related to retest reliability. This suggests the possibility of classifying items according to probable stability on the basis of these variables. As used in the present study, a classifier is a measure used to separate the components of a predictor(s)—in this case items in a biographical inventory—into

categories of expected reliability. Classifier variables might provide a means of improving the predictive validity of life history forms on which a scoring system must be derived concurrently. In addition, given a large fixed item set serving as a multi-criteria/multi-population prediction device, classifiers could conceivably prove valuable in sorting out the items according to their probable stability in each situational context.

Starry (1966), working with discrete data (binary scoring of each alternative), developed a biographical item classification system based on social desirability indexes (Edwards, 1957) for the prediction of item reliability and fakeability. Correlations of .31 and .45 were obtained between social desirability ratings and test-retest stability and fakeability response measures, respectively. Considerable shrinkage occurred when prediction weights developed on the 100-item experimental questionnaire were cross-validated on an equivalent form. A follow-up study conducted by Starry and Tesser (1967) using only those items in the questionnaires significantly related to a grade-point criterion produced similar results. The utility of this approach seemed to be limited to the prediction of item fakeability, and then only as a relative measure within a particular item sample.

¹ Requests for reprints should be sent to A. R. Starry, Director, Measurement and Research Center, Purdue University, Lafayette, Indiana 47907.

The social desirability scale used in these previous studies generalizes well to most discrete items, but loses relevance when applied to continuously scored biographical items or to cases in which one is interested in viewing a set of qualitative alternatives as a whole. It is meaningful to describe the social desirability of a particular response alternative, but not the social desirability of a set of qualitative or continuous alternatives. One could, however, talk about whether the set of alternatives as a whole would be more or less prone to social desirability response bias. In this case we would be dealing with the variance of the alternatives with respect to their social desirability values. One might argue that response stability is related in some way to the degree to which a set of response alternatives are varied on this dimension.

The problem here, as indeed the problem with our earlier work (Starry, 1966; Starry & Tesser, 1967), lies in the fact that it may not go far enough. In the first place two different approaches are available—one for a particular, discretely scored response alternative and one for a set of alternatives. Second, it is possible that social desirability ratings *per se* may be too general for efficient prediction in some particular applied setting. A set of response alternatives may in general have a large social desirability component which may be totally irrelevant for the particular purpose for which the respondent is taking the instrument. For example, it may be more socially desirable, on a superficial level, for a respondent to be able to play some musical instrument than to have no ability in this area, but this same item could be totally irrelevant if the respondent is taking the inventory in conjunction with a job application as a machinist. Or, with the system used in the discrete case, it is possible that what is generally socially desirable is not socially desirable for a particular setting or vice versa. To appear gregarious, in general socially desirable, may be an unacceptable kind of response when one is applying to become an astronaut.

There are many other sources of variance that affect response stability besides respondent attempts to make the appropriate response in a particular setting. For example, an item may ask for information that a

respondent doesn't have, it may make finer discriminations than the respondent can handle, it may ask for information that is dimly remembered, or it may be sensitive to the respondent's mood at a particular time, etc.

A scale called Probable Response Stability (PRS) was developed in an attempt to find a straightforward measure which would be sensitive to the issues discussed above. That is, it should be useful for looking at discrete items as well as sets of alternatives. In addition it should reflect not only differences in situation-specific social desirability but other sources of response stability variance as well.

The assumption underlying this scale is that external raters armed with information about the situational context, purpose of the instrument, and respondent population can make judgments about how stable the responses to the items will be. If this rationale has any merit one would expect two outcomes: First, that the judgments made by raters will be reliable (indicating that there is something in the task that can be uncovered) and second that the PRS scale will make a significant contribution to the prediction of response stability.

The investigations reported here are attempts to evaluate the PRS scale as a potential classifier variable. Methodological considerations in the use of classifiers and empirical evaluations of their usefulness in personal history research must be the subject of future explorations. Two sets of data were used in the following studies: one set based upon discrete (binary) scoring of biographical data and the other on continuously scored items. The analysis of discrete items will be referred to as Study I and the analysis of continuous items as Study II.

STUDY I

Method

The method used to collect reliability and fakeability measures on discrete items has been described previously in detail (Starry, 1966; Starry & Tesser, 1967). In summary, two 100-item questionnaires were administered to a sample of 321 undergraduate males. Half of Ss who received Form A and half of Ss who received Form B were instructed to respond to the items in such a way as to make

themselves appear maximally attractive to persons using this instrument as a basis for admitting students into graduate school (fake set). The remaining Ss were asked to respond honestly to the items (honest set). Approximately 3 mo. later, the instruments were readministered to all Ss under the honest set. Item analysis was conducted on both questionnaire forms using a criterion of college grade point average, resulting in 60 discriminating items for Form A and 52 for Form B. The item stability measures (i.e., reliability and fakeability) were generated by calculating the percentage of consistent responses to each item's discriminating alternative on the two administrations.

To obtain the item classification information for the present study, discriminating alternatives in both biographical forms were rated by 26 advanced undergraduates with the PRS scale shown below. Judges were instructed to read the entire questionnaire before assigning ratings and to consider each designated alternative within the complete item context.

| Probable Response Stability Scale | | | | | | | | |
|--|---|---|--------------------|---|---|----------------|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Low Stability | | | Moderate Stability | | | High Stability | | |
| | | | | | | | | |
| The rating task is to judge, on this 9-point scale, the extent to which designated responses to certain items on the questionnaire are likely to remain stable over a 3-mo. time interval. | | | | | | | | |

They were also informed that the respondent population consisted of male undergraduates. A sample rating with a verbal description of the logic which had been used in arriving at a given stability value was included in the instructions. Mean stability ratings of the 112 discriminating alternatives ranged from 4.58 to 8.35.

Results

Estimated reliability of the mean ratings was .85 for Form A and .86 for Form B

TABLE 1
CORRELATIONS BETWEEN MEAN PRS RATINGS
AND ITEM STABILITY INDEXES FOR
DISCRETELY SCORED ITEMS

| Item index | Probable response stability ratings | |
|-------------|-------------------------------------|----------------------|
| | Form A (60 items) | Form B (52 items) |
| Reliability | .54* | .51* |
| Fakeability | .53* | .50* |

* $p < .01$.

items (Winer, 1962, p. 126). These results are summarized in Table 1. Validity coefficients for the ratings were in the low .50's, indicating a moderate degree of overlap between PRS and actual stability and fakeability.

Total score reliability (test-retest) dropped from .79 to .50 (Form A) and from .78 to .47 (Form B) when the fake set was introduced. There was some tendency for those items on which faking took place to be less reliable. Stability of items under honest test-retest conditions correlated .63 and .67 with that obtained under the influence of this fake set.

STUDY II

Method

During orientation week in September, 1967, Purdue freshmen completed a 671-item biographical questionnaire constructed for some other research purposes. From this instrument the authors selected 88 five-alternative, continuously scored items representative of content areas sampled by the complete questionnaire. Although question topics overlapped in some areas, none of the items had appeared in the questionnaire used in Study I. These items were then readministered 16–20 wk. later to a volunteer group of 106 male and female freshmen. The Ss were instructed to respond as accurately as possible, with no fake set being introduced.

A stability index for each item was obtained by computing the absolute difference between responses given by each S on the two administrations, summing across all Ss, and dividing by the number responding to that item on both administrations (N per item varied from 104 to 106). In order to express the stability index and PRS ratings in more similar terms, this quantity was then subtracted from a constant of 9.00. The stability index ranged from 7.86 for the last stable item in this questionnaire to 8.89 for the most stable.

The 88 items were rated for stability by 25 male and female graduate students with the PRS scale used in Study I. Instructions were modified to accommodate the different item type. Judges were instructed to assign probable stability values to the total item instead of to a single designated item alternative as was the case with discrete scoring. Mean PRS ranged from 5.44 to 9.00.

Results

Reliability of the mean PRS ratings, computed as before, was estimated to be .97. The product moment correlation between item stability and mean PRS values across the 88 items was .47, significant at the .01 level. The correlation between mean rating and item stability index standard deviation was

calculated as a check on the interpretation of the rating task. The lack of association between these variables ($r = -.03$) tends to indicate that judges were not attempting to rate the probable stability of responses between (as opposed to within) the respondents.

DISCUSSION

The PRS rating scale seems to offer the universality necessary for rating both item types with acceptable interrater reliability, although its effective range was less than four scale points in both cases.

Little practical difference was found in these studies between the predictability of test-retest item stability for discrete or continuous type biographical items. Although the fakeability and stability indexes for the same discrete items were rather dissimilar, predictability was virtually the same for both. External judges were able to account for approximately 25% of the criterion variance in either item type.

While stability ratings would seem to offer some practical utility as a classifier variable in applied prediction studies with biographical items, several problems must first be overcome. Perhaps the most serious concerns the shape of the bivariate distribution of rated (PRS scale) versus empirical stability. Although the scatter-plots produced in these studies displayed a fair degree of linear regression, homoscedasticity was weak. In particular, prediction appeared best at the "High Stable" end of the scale. Unless this phenomenon is merely sampling fluctuation, elimination of some of the lowest rated items in a questionnaire would not result in the increase to total score stability which coefficients around .50 would suggest. To the researcher who could afford it, deletion of a large percentage of the lower rated items would probably increase the average stability of remaining items, but total score stability may not improve because of the effects of shortening the instrument. Increased emphasis on the attainment of normal distributions during construction of classifiers and the biographical instruments themselves (Starry, 1968) seems advisable. In addition, the relationship between item reliability and validity is almost certain to be partially dependent on the

criterion under investigation. It is conceivable that in certain prediction settings items judged to be relatively less stable could actually be the most valid. Their elimination might seriously impair the validity of the biographical form under investigation. The solution to both these problems would appear to be some effective system for the prediction of stability which takes item validity into account and gives the researcher more assurance that the items he deletes because of low rated stability are actually too unstable for inclusion in his selection instrument. Perhaps a combination of classifier information used in conjunction with routine concurrent item analysis statistics will provide such a system.

The authors have accumulated some evidence in pilot studies conducted during development of the PRS scale that rater knowledge of respondent and criterion characteristics does not enhance their ability to predict item stability, suggesting some degree of scale invariance. However, the extent to which item ratings might be invariant with respect to the apparent range of stability peculiar to the set of items of which they are a part is unknown. An item rated as moderately stable in one context might be rated highly stable in another. Until this matter is resolved and rated values associated with predicted ranges of empirical stability, the researcher working with a particular item set will find it difficult to formulate standards and cutoff points for item deletion.

Total score test-retest reliability coefficients in the high .70's for the 52 and 60 item discretely scored questionnaires, while perhaps not as large as might be desired, are reasonable in lieu of the uncontrollable sources of variation encountered with "volunteer" college student respondents. It is this lack of control which may account for the low reliability of .60 obtained with the 88-item continuously scored questionnaire. Although every effort was made to elicit accurate information on both administrations, the length of the original questionnaire of which these items were a part undoubtedly resulted in carelessness on the part of some students. Low reliability in Study II might also be related to the specific respondent sample which was com-

posed of first semester college freshmen, a group highly subject to changes in value systems and perceptions. All the questionnaire reliability coefficients reported here should probably be viewed as lower-bound estimates.

REFERENCES

- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- KLEIN, S. P., & OWENS, W. A., JR. Faking of a scored life history blank as a function of criterion objectivity. *Journal of Applied Psychology*, 1965, 6, 452-454.
- LARSEN, R. H., SWARTHOUT, D. M., & WICKERT, F. R. *Objectionability and fakeability of biographical inventory items*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 7, 1967.
- OWENS, W. A., JR., GLENNON, J. R., & ALBRIGHT, L. E. Retest consistency and the writing of life history items: A first step. *Journal of Applied Psychology*, 1962, 5, 329-331.
- STARRY, A. R. The stability issue in biographical prediction. *Studies in Higher Education*, 1966, 93, 19-28.
- STARRY, A. R. *The utility of biographical data for screening research and development engineering applicants*. Symposium paper presented at the meeting of the American Psychological Association, San Francisco, September 1968.
- STARRY, A. R., & TESSER, A. *Social desirability ratings as predictors of biographical item stability and fakeability*. Paper presented at the meeting of the Midwestern Psychological Association, May 1967.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Early publication received July 17, 1968)

SPEED AND PRACTICE: EFFECTS ON NEGRO AND WHITE TEST PERFORMANCES¹

JERRY A. DUBIN² AND HOBART OSBURN

University of Houston

AND DARVIN M. WINICK

Lifson, Wilson, Ferguson & Winick, Inc., Houston, Texas

This research investigated the hypotheses that (a) extra pretest practice, (b) extra testing time, and (c) extra practice and extra testing time would improve the mental ability test performances of Negroes more than whites. The Ss, Negro and white high school students in the higher and lower socioeconomic classes, were administered parallel forms of several ability tests. Some Ss took the tests under speeded conditions, others under power conditions. Although both races and both socioeconomic classes improved their performances as the testing procedures became more lenient, all groups profited to a comparable extent; the three hypotheses were rejected. Implications are that the testing procedure itself does not discriminate between racial groups nor between culturally advantaged and disadvantaged Ss.

The question of whether ability tests unfairly discriminate against minority groups is of great concern to psychology, education, and industry. Researchers interested in this problem have generally focused their attention on two aspects of discrimination, that is, test content and analysis of test results. Another potential aspect of unfair discrimination involves the testing procedure itself. Since it is conceivable that certain testing conditions systematically favor one cultural group over another, variables such as test administrator's race, test directions, methods of responding, testing time, and amount of pretest practice need more attention.

The purpose of the present study was to determine if highly speeded tests are equally fair to Negroes and whites and, as other research indicates (Boger, 1952; Eagleson, 1937; Katzenmeyer, 1962; Klineberg, 1928; Vane & Kessler, 1964), extra practice or test familiarity would reduce the Negro-white test score discrepancies. It was hypothesized that

Negroes would benefit more than whites when opportunities were available for (a) extra pretest practice, (b) extra testing time, and (c) both extra practice and extra testing time.

METHOD

Subjects

Two hundred and thirty-five Negro students from a predominately Negro high school and a random sample of 232 white students from a predominately white high school in the same school system served as Ss. The Negro and white students were divided into four groups: S1, S2, P1, and P2. Groups S were administered speeded tests, whereas Groups P received power tests; Groups 1 were comprised of students in the ninth and tenth grades, while Groups 2 were eleventh and twelfth graders. Since both schools were within the progressive Galena Park (Texas) School District, all Ss were quite familiar with standardized tests.

Ability Tests

Forms A and B of four Employee Aptitude Survey (EAS) tests were used: Numerical Ability (solving computational problems), Space Visualization (counting three-dimensional blocks), Numerical Reasoning (solving number series items), and Verbal Reasoning (drawing valid conclusions from a list of facts). Both forms for each test were constructed in a parallel manner and are statistically equivalent (Ruch and Ruch, 1963).

Procedure

Before testing began, Ss were instructed that each test had two parts (actually the alternate forms)

¹ This study is based on a thesis submitted by the senior author to the Department of Psychology, University of Houston, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

² Now with Lifson, Wilson, Ferguson & Winick, Inc., Management Consultants, 1811 Crawford, Houston, Texas 77002. Requests for reprints should be sent to Jerry D. Dubin at the above address.

TABLE 1
DISTRIBUTION OF THE SAMPLE BY GRADE LEVEL,
RACE, SOCIOECONOMIC STATUS, AND TESTING
CONDITIONS

| Ss | Sample size | | |
|-------------------|------------------------------|--------------------------------|------------------|
| | Timed admini- stration | Untimed adminis- tration | Matched pairs |
| 9th-10th graders | | | |
| Negro | | | |
| Low SES | 24 | 29 | 14 |
| High SES | 33 | 18 | 14 |
| White | | | |
| Low SES | 28 | 32 | 25 |
| High SES | 29 | 26 | 25 |
| Total | 114 | 105 | 78 |
| 11th-12th graders | | | |
| Negro | | | |
| Low SES | 44 | 38 | 32 |
| High SES | 26 | 23 | 10 |
| White | | | |
| Low SES | 31 | 33 | 31 |
| High SES | 27 | 26 | 24 |
| Total | 128 | 120 | 97 |

to be taken consecutively. Negroes and whites in Group S1 were administered both forms of the "speeded" Numerical Reasoning and Space Visualization Tests with the regular 5-min. time limits; Group P1 was administered both forms of the "power" Numerical Reasoning Test with tripled time limits. Similarly, Group S2 was administered both forms of the Verbal Reasoning and Numerical Ability Tests with the regular 5- and 10-min. time limits and Group P2 was administered both forms of the Verbal Reasoning Test with tripled time limits. In an attempt to reduce the effects of the test administrator's race (all tests were administered by whites), Negro teachers assisted in the testing of Negro students. Upon finishing the tests Ss were asked to complete a short socioeconomic questionnaire. Based on four items (father's occupation, father's education, mother's education, and student's educational expectancies), a socioeconomic status (SES) index was derived. Dichotomizing the index for both racial groups resulted in four categories: high SES whites (HW); low SES whites (LW), high SES Negroes (HN), and low SES Negroes (LN). See Columns 1 and 2 in Table 1 for sample distributions. The tests and questionnaires were administered during a 2-day session in December 1967.

RESULTS

Raw Test Scores

Figure 1 illustrates a relationship between test performance and racial-socioeconomic fac-

tors. Typically, on each test for each condition, the whites outperformed the Negroes, and the culturally advantaged students outperformed the culturally disadvantaged (Table 2). With one minor exception, an order was maintained; the high SES whites performed best followed by the LW, HN, and LN groups. (Somewhat similar results were found by Fifer, 1964.)

In another expected finding, Figure 1 demonstrates that all Ss attained higher test scores as the testing situation became more lenient. Mean test scores increased consistently as the procedure progressed from speeded tests-no practice, to speeded tests-practice, to power tests-no practice, to power tests-practice. This steady improvement was found for both racial groups and both SES groups. The improvements obtained by Negroes and whites are compared below, separately for each hypothesis.

Hypothesis 1

In analyzing the hypothesis that extra practice would be more advantageous to Negro than to white Ss, the differences between Forms A and B for each test were analyzed by the double classification analysis of variance model.⁸ In six investigations of practice effects there were no significant differences attributable to race (Table 3). Consequently, the extra practice equally enhanced Negro and white test performances, and Hypothesis 1 was clearly refuted.

Furthermore, Table 3 shows that extra practice profits both socioeconomic classes to a comparable degree. (The one exception is probably due to a sampling error, see Figure 1a.) Unexpectedly, the Race \times SES interaction was significant for two of the four speeded tests. This suggests that the low SES whites and high SES Negroes improve most when extra practice is provided on certain types of speeded tests.

Hypothesis 2

To investigate the hypothesis that extra testing time would favor Negro students, Ss taking speeded tests (Groups S1 or S2) were

⁸ Because the cell sizes were unequal the sum of squares for each analysis was estimated by the least-squares solution (see Winer, 1962).

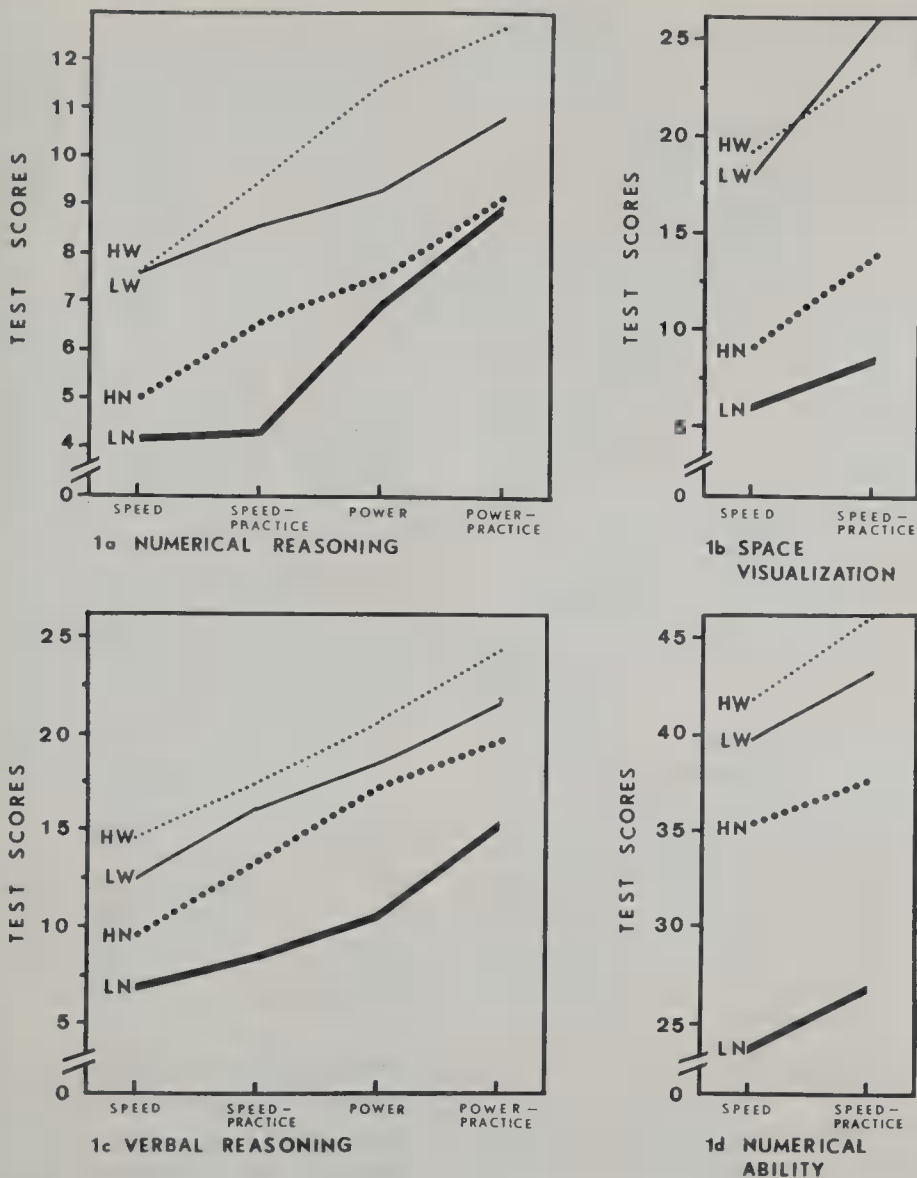


FIG. 1. Average score for each test by race, socioeconomic status, and testing conditions.

paired with Ss taking power tests (Groups P1 or P2). Pairings were controlled by matching Ss for race, sex, grade level, and SES index. (Because of this restrictive matching procedure, the sample was substantially reduced, see Column 3 in Table 1.) Hypothesis 2 was then tested by comparing the speed-power test differences between the matched Negroes with the differences between the matched whites.

As shown in Table 3, Negro and white improvement scores do not significantly differ. Hypothesis 2 was therefore rejected. Similarly, socioeconomic status and the Race \times SES interaction were unrelated to improvements resulting from extra testing time.

Consequently, administering tests without time limits does not favor Negroes or culturally disadvantaged students.

Hypothesis 3

The third hypothesis, that Negroes will benefit most from the combination of extra practice and extra time, also was tested by matching Ss. Using the same matched pairs as used for testing the previous hypothesis, the difference scores (in this case Form B untimed minus Form A timed) were calculated and analyzed. Again, the hypothesis was rejected. Neither race, socioeconomic status, nor the Race \times SES interaction was significantly related to score improvements resulting

from extra practice and extra testing time (Table 3).

DISCUSSION AND CONCLUSIONS

The results failed to support the hypotheses that Negroes would be favored by extra practice and/or extra testing time. Apparently, the administration of highly speeded tests given without extra practice did not handicap the average Negro S nor Ss in lower socio-economic classes. In a more general sense the results imply that the testing procedure itself was not a major factor in discriminating between Negro and white Ss or between culturally advantaged and culturally disadvantaged Ss.

Although the main effects of race and SES were not significant, the Race \times SES interaction was significant for two of the four timed tests. On these tests, Verbal Reasoning and Space Visualization, the low SES whites and high SES Negroes improved more when given extra practice than did the high SES whites and low SES Negroes. Neither of the unspeeded tests showed a significant interaction effect. Why the low SES whites and high SES Negroes benefited more from extra practice and why the interactions were significant only for particular tests and only under speeded conditions is open to speculation.

One possible explanation is that Ss in the low SES white and high SES Negro groups do not function near their capacity in re-

TABLE 2
SUMMARY OF F RATIOS FOR TWO-WAY ANALYSIS OF VARIANCES USING TEST SCORE DATA

| Test/treatment | df | Source of variance | | |
|---------------------|-------|--------------------|----------|-------------------|
| | | Race | SES | Race \times SES |
| Numerical reasoning | | | | |
| Timed A | 1/110 | 20.638** | 0.374 | 0.325 |
| Timed B | 1/110 | 28.422** | 5.832* | 0.876 |
| Untimed A | 1/101 | 14.371** | 3.728 | 1.001 |
| Untimed B | 1/101 | 9.499** | 1.897 | 0.957 |
| Space visualization | | | | |
| Timed A | 1/110 | 49.783** | 1.790 | 0.335 |
| Timed B | 1/110 | 38.894** | 0.560 | 3.728 |
| Verbal reasoning | | | | |
| Timed A | 1/124 | 23.043** | 5.230* | 0.076 |
| Timed B | 1/124 | 30.435** | 8.670** | 2.646 |
| Untimed A | 1/116 | 23.649** | 12.980** | 3.069 |
| Untimed B | 1/116 | 24.548** | 9.764** | 0.435 |
| Numerical ability | | | | |
| Timed A | 1/124 | 25.834** | 9.531** | 3.938* |
| Timed B | 1/124 | 27.004** | 7.383** | 2.391 |

* $p < .05$.
** $p < .01$.

TABLE 3
ANALYSIS OF VARIANCE SUMMARIES FOR DIFFERENCE SCORES

| Tests ^a | df | Source of variance | | |
|--|-------|--------------------|--------|------------|
| | | Race | SES | Race × SES |
| Hypothesis 1: Extra practice (Form B minus Form A) | | | | |
| NR speeded | 1/110 | 1.313 | 6.153* | 0.258 |
| SV speeded | 1/110 | 2.218 | 0.950 | 6.792* |
| VR speeded | 1/124 | 1.236 | 1.059 | 4.350* |
| NA speeded | 1/124 | 1.501 | 0.285 | 0.852 |
| NR power | 1/101 | 0.569 | 0.393 | 0.000 |
| VR power | 1/116 | 0.304 | 1.185 | 2.684 |
| Hypothesis 2: Extra time (untimed minus timed) | | | | |
| NR Form A | 1/74 | 2.601 | 1.385 | 0.786 |
| NR Form B | 1/74 | 0.058 | 0.047 | 0.361 |
| VR Form A | 1/93 | 0.437 | 0.160 | 0.516 |
| VR Form B | 1/93 | 0.504 | 0.153 | 0.499 |
| Hypothesis 3: Extra practice and time (untimed Form B minus timed Form A) | | | | |
| NR | 1/74 | 1.005 | 0.658 | 0.198 |
| VR | 1/93 | 0.009 | 0.037 | 0.003 |

^a NR = Numerical Reasoning; SV = Space Visualization; VR = Verbal Reasoning; NA = Numerical Ability.
* $p < .05$.

stricted situations, that is, when examined on uncommon tasks⁴ or tested under highly speeded conditions. When provided with extra pretest practice, therefore, these Ss work more effectively and increase their test scores substantially. The low SES Negroes, the most culturally deprived Ss, are also functioning below their potential. Their skills or test-taking abilities, however, are so underdeveloped that a single practice test is not sufficient to allow them to master the unfamiliar tasks and the highly speeded conditions. On the other hand, the most culturally advantaged Ss (i.e., the high SES whites) have broader experiences and, hence, are better prepared for the tasks. Consequently, they work rather well on the first administration and only improve to a moderate extent when provided with extra practice.

⁴ Because of the school district's testing program, all students were quite familiar with the two numerical tests but unfamiliar with the spatial and verbal tests.

The findings of this research must be considered with respect to the study's limitations. One of the major limitations was the need to match Ss taking speeded and power tests. The insignificant results for Hypotheses 2 and 3, therefore, could be related to the matching procedure. A second limitation is the fact that Ss were a homogeneous group of high school students experienced with standardized testing. It is not known if the results are generalizable to Ss less experienced in test taking, to older Ss, or to industrial populations. A third limitation was the use of only one practice test. Remaining to be tested are the effects of a series of practice tests.

With the above limitations noted, the most important implication of this research is that speeded tests do not handicap Negro Ss nor are they likely to handicap the future Negro job applicant.

REFERENCES

- BOGER, J. H. An experimental study of the effects of perceptual training on group IQ test scores of elementary pupils in rural ungraded schools. *Journal of Educational Research*, 1952, 46, 42-52.
- EAGLESON, O. W. Comparative studies of white and Negro subjects in learning to discriminate visual magnitude. *Journal of Psychology*, 1937, 4, 167-197.
- FIFER, G. Social class and cultural group differences in diverse mental abilities. *Proceedings of the invitational conference on testing problems*. Princeton: Educational Testing Service, 1964, 107-117.
- KATZENMEYER, W. G. Social interaction and differences in intelligence test performance of Negro and white elementary school pupils. Unpublished doctoral dissertation, Duke University, 1962.
- KLINEBERG, O. An experimental study of speed and other factors in racial differences. *Archives of Psychology*, 1928, 15, 109-122.
- RUCH, F. L., & RUCH, W. W. *Technical report: Employee aptitude survey*. Los Angeles: Psychological Services, Inc., 1963.
- VANE, J. R., & KESSLER, R. T. The Goodenough Draw-A-Man Test: Long term reliability and validity. *Journal of Clinical Psychology*, 1964, 20, 487-488.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Early publication received August 19, 1968)

CONTRIBUTIONS OF THE INTERVIEW TO ASSESSMENT OF MANAGEMENT POTENTIAL

DONALD L. GRANT¹ AND DOUGLAS W. BRAY

American Telephone and Telegraph Company

The contribution of interview information to assessment center evaluations and the relationship of interview variables to progress in management are presented. The interview data were obtained by coding interview reports. Analyses of the data clearly indicate that information from the interview reports contributes to assessment center evaluations. Judgments of career motivation and to a lesser extent work motivation and control of feelings appear to have been influenced by the interview information. In addition, judgments of interpersonal skills were reinforced, if not influenced, by the interview reports. The results of the study also demonstrate that extensive and reliable information on many personal characteristics can be obtained from the interview. In addition, several of the interview variables, especially those reflecting career motivation, dependency needs, work motivation, and interpersonal skills are directly related to progress in management. The findings clearly indicate that relevant information on personal characteristics important to managerial success was obtained from interview reports.

The Bell System Management Progress Study, a longitudinal investigation of the development of young men in a business management environment (Bray, 1964), provides the opportunity for a thorough investigation of assessment center procedures. An important part of such an investigation is a study of the contributions of the various assessment techniques to assessment staff judgments. The center used in the study has been described and the results of analyses for most of the assessment techniques presented (Bray & Grant, 1966; Grant, Katkovsky & Bray, 1967). In addition to examining the relationships of the techniques to assessment staff judgments, correlations of each technique and of staff judgments to a progress criterion have been reported.

A major omission from the previous analyses has been the assessment interview. The present article presents information on the contributions made by the interview to the assessment center process and relationships of interview variables to the progress criterion.

The interview is, of course, the most widely used method of evaluating candidates for employment, including candidates for manage-

ment positions. Research evidence in support of assessing people by means of the interview, however, is limited. Two recent reviews of research on the selection interview (Mayfield, 1964; Ulrich & Trumbo, 1965) raise questions concerning the technique. Mayfield (1964) notes that data supporting the selection interview are not substantial. He particularly questions the consistency of materials covered and inter-rater reliability in the unstructured interview. He concludes, moreover, that even where the reliabilities of the selection interview, structured or unstructured, are high, "the validities obtained are usually of a low magnitude" (Mayfield, 1964, p. 251). In addition, he declares that the only characteristic which can be estimated reliably and validly from interviews is that of mental ability.

Ulrich and Trumbo (1965) also found data favoring structured over unstructured interviews in selection. The structured interviews have proved more valid. In contrast to Mayfield, however, they conclude that the interviewer can most validly assess the areas of personal relations and motivation to work (Ulrich & Trumbo, 1965, p. 113).

In addition to its widespread use in personnel selection the interview has been an integral part of assessment center procedures. In general, however, its contributions to the assess-

¹ Requests for reprints should be sent to Donald L. Grant, Personnel Manager, Research, American Telephone and Telegraph Company, 195 Broadway, Room 2122, New York, New York 10007.

ment process, along with those of other assessment center techniques, have not been studied. A partial exception to this generalization is the assessment center activity of the Office of Strategic Services in World War II (OSS Assessment Staff, 1948). In their centers the interview was considered the most important single procedure and the interviewer's rating had the most influence on the final decision by the assessment staff concerning a candidate's acceptability. The interviewer's rating, however, was not based solely on the interview. Prior to the interview the interviewer was furnished with considerable information on the candidate (personal history record, health inventory, projective questionnaire, etc.). He also had the opportunity to observe the candidate in many situations. As a consequence, the impact of the interview *per se* could not be ascertained.

The predictive value of the interview when used in assessment center or analogous activities has been reported by a few investigators. Kelly and Fiske (1951) report that the interview added little to the prediction of subsequent performance by clinical psychologists. MacKinnon (1958) presents correlations between interviewer ratings and various criteria of the effectiveness of Air Force officers. Many are statistically insignificant and the remainder low (i.e., in the .20s). In a relatively detailed report Prien (1962) presents data showing rather low reliabilities between interviewers' ratings, statistically insignificant correlations between interviewers' ratings and supervisors' ratings of the performance of sales candidates, and several statistically reliable correlations between the interviewers' ratings and supervisors' ratings of performance of candidates for managerial and technical positions.

The data from the Bell System Management Progress Study assessment center interviews have been examined to ascertain their relationships to assessment staff judgments which have been shown to be predictive of progress in management; and the variables from the interviews have been related to a progress criterion. Neither the interview data nor any other assessment information have been available to Bell System management or

to Ss. There is, therefore, no contamination of the progress criterion.

METHOD

Interviewing

The interviews in the Management Progress Study assessment centers are relatively unstructured. Prior to the interview the interviewer is furnished with a completed personal history record on *S* which he reviews for essential biographical information and for areas in which to probe during the interview (e.g., relationships with parents and siblings). The interviewer also is instructed to cover a number of topics during the interview such as work goals, attitudes on social issues, and hobbies.

The interview is conducted so as to insure privacy. Each *S* is asked by the interviewer for permission to take notes and is assured of the confidentiality of anything he says. None of the interviews has been recorded. The interviews are relatively informal, the interviewer being free to follow leads as they develop in the interview and to vary his style of interviewing (directive or nondirective) in accord with circumstances.

The assessment schedule allots 2 hr. to the interview, though it may be terminated in less than the scheduled time if in the judgment of the interviewer all pertinent topics have been covered adequately. Upon completion of the interview *S* is reassured of the confidentiality of the information, a report to be made to the assessment staff only.

Immediately following each interview the interviewer, using a dictating machine, dictates a report from his notes. He is not asked to rate *S* at this time. The recorded interview report is played to the assessment staff at the evaluation meeting. At this time the interviewer, as a member of the staff, participates in evaluating *S* after the evidence from all the assessment techniques have been reported (Bray & Grant, 1966). At a later date the interview report is transcribed and filed with all the assessment data concerning *S*.

In general the assessment center interviewers are professional psychologists. The interviews analyzed for this article were conducted by six persons, five of whom are psychologists.² A majority of the interviews were conducted by two of the five psychologists.

Analysis

The Ss for this investigation of the interview are 348 men who are participating in the Management Progress Study. Of these, 200 had graduated from college prior to Bell System employment while 148 had not been employed as college graduates but had been promoted to management early in their careers. The great majority of Ss were in their 20s when

² The interviews were conducted by Warren D. Bachelis, David E. Berlew, Donald C. Dewar, John J. Hopkins, David B. Muirhead, and Joseph F. Rychlak.

originally assessed; a few of the noncollege Ss were in their early 30s. The Ss came from several regions of the country.

For purposes of analysis each interview report was coded independently by two advanced graduate students majoring in psychology who had not participated in the assessment.³ Using a manual the coders rated each report on 18 variables. The variables were selected from the 25 variables rated by the assessment center staffs in making their evaluations of Ss (Bray & Grant, 1966). After reviewing several interview reports it was judged likely that information pertinent to these variables had been obtained by the interviewers. The variables and their definitions are

Personal Impact-Forcefulness

How forceful an early impression does he make? Consider the impression he made on the interviewer.

Oral Communications Skills

How effectively does he express himself? Consider ease of expression, correct use of English, vocabulary, precision in explaining views, vocal clarity, and tonal quality.

Human Relations Skills

How well can this man get people to perform effectively by good human relations techniques? (His sincerity is irrelevant.)

Personal Impact-Likability

How likable an early impression does he make? Consider the impression he made on the interviewer. Did the interviewer tend to like or dislike him?

Behavior Flexibility

How readily can he, when motivated, modify his behavior to reach a goal? Consider tendencies to persevere and frequency with which he has adapted to changing circumstances.

Need Approval of Superiors

To what extent does he seek approval of persons in authority over him? Consider his dependence on superordinates for help and guidance as well as tendencies to solicit praise and support from them.

Need Approval of Peers

To what extent does he seek approval of his peers? Consider his dependence on his coordinates for help and guidance as well as tendencies to solicit support from them.

Tolerance of Uncertainty

To what extent will his work performance stand up under uncertain or unstructured conditions? Consider his need for structure and the impact of lack of structure on his behavior.

Inner Work Standards

To what extent will he want to do a good job even if a less good one is acceptable to his boss and others? Consider the quality of results he expects of himself and of others (e.g., subordinates).

Primacy of Work

To what extent will he find satisfactions from work more important than satisfactions from other areas of life? Consider the value he places on work, the satisfactions he obtains from it relative to other satisfactions (e.g., family, hobbies, community activities) and his willingness to devote more than the required time to his job.

Energy

How continuously can he sustain a high level of work activity? Consider his general activity level, the effort he puts into his work, and his reactions to expending energy (e.g., evidence of fatigue).

Goal Flexibility

To what extent will he be able to change his life goals (such as money, power, fame, etc.) in accordance with reality opportunities? Consider what he says are his goals and his commitment to them.

Need Advancement

To what extent will he need to be promoted significantly earlier than his peers in a job? Consider the level he aspires to and the rapidity with which he expects to achieve it.

Need Security

To what extent does this man need a secure job (not necessarily with the Bell System)? Consider his motives in accepting a position in the Bell System, his views about leaving the System, and his views regarding alternative employment.

Social Objectivity

How free is he from prejudices against racial, ethnic, socioeconomic, educational, and other kinds of groups? Consider strength and inclusiveness of prejudice.

Bell System Value Orientation

To what extent is he likely to incorporate early in his career Bell System values such as service, friendliness, justice of company position on rate increases, etc. Consider his identification with the System, including his desire to remain in it despite possible disappointment of his personal goals.

Ability to Delay Gratification

To what extent will this man be able to work over long periods of time without great rewards in order to reach later rewards? Consider his tolerance for frustration, patience, and the long-range vs. short-range nature of his goals.

Range of Interests

To what extent is he interested in a variety of fields of human activity such as science, politics, sports, music, art, etc? Consider his leisure time activities, hobbies, reading habits, community activities, etc.

A 5-point scale was used in rating each variable. As an example the scale for Oral Communications Skills is shown below:

1. Expresses himself very poorly.
2. Expresses himself rather poorly.
3. Expresses himself well in some ways, poorly in others.

³ The interview reports were coded by Byron Fiman and Virginia Ellen Schein.

4. Expresses himself well.
5. Expresses himself very well.

The coders were instructed to read each interview report and to note, using abbreviations for the variables in the margin, information pertinent to evaluating the variables. After reviewing the evidence for a given variable the coders, using the scale provided, recorded a rating. The coders were further instructed to omit ratings on variables for which the evidence was inadequate and, as far as possible, to avoid the error of central tendency in making their ratings.

The first step in analyzing the data was to group the coded reports according to the educational backgrounds of Ss (i.e., college graduates and noncollege) at time of employment. For each variable the percentage of reports rated by both coders in each sample was computed. The reliability of the coding procedure was determined by correlating the ratings of the two coders (omitting ratings not recorded by both) on each variable for each sample of Ss. The Spearman-Brown prophecy formula then was applied to the correlations.

The ratings of the coders (using only those recorded by both) were pooled by simple addition and from the resulting sum scores the means, standard deviations, variances, and intercorrelations of the variables computed. The sum scores also were correlated with the judgments of the assessment staff, scores from other assessment techniques, and progress in management as reflected by a salary criterion. In essence, the analyses made are parallel to those for the other assessment techniques previously studied (Bray & Grant, 1966; Grant, Katkovsky, & Bray, 1967).

RESULTS

Table 1 shows for each variable the number and percentage of interviews which both coders were able to rate. The percentages for the college graduate sample range from 56% to 99%, averaging 90% with 12 of the 18 variables being rated for 90% or more of the interviews. The corresponding figures for the noncollege group are markedly lower. They range from 33% to 99%, average 79%, and only 8 of the variables have percentages of 90 or better. No explanation is at hand to explain the lesser ratable of the noncollege interviews. It could be that interviewers and coders with extensive higher education are more effective in evaluating college-educated Ss.

Several variables were nearly always ratable for both samples. These included behavior flexibility, goal flexibility, need advancement, Bell System value orientation, ability to delay gratification, and range of

TABLE 1
NUMBER AND PERCENTAGE OF INTERVIEWS CODED

| Variable | College sample (N = 200) | | Noncollege sample (N = 148) | |
|--------------------------------|-----------------------------|----|--------------------------------|----|
| | N | % | N | % |
| Personal Impact-Forcefulness | 157 | 79 | 49 | 33 |
| Oral Communication Skills | 155 | 78 | 95 | 64 |
| Human Relations Skills | 185 | 93 | 56 | 38 |
| Personal Impact-Likability | 111 | 56 | 51 | 34 |
| Behavior Flexibility | 198 | 99 | 143 | 97 |
| Need Approval-Superiors | 165 | 83 | 116 | 78 |
| Need Approval-Peers | 170 | 85 | 105 | 71 |
| Tolerance of Uncertainty | 191 | 96 | 125 | 84 |
| Inner Work Standards | 197 | 99 | 135 | 91 |
| Primacy of Work | 185 | 93 | 142 | 96 |
| Energy | 187 | 94 | 114 | 77 |
| Goal Flexibility | 190 | 95 | 146 | 99 |
| Need Advancement | 197 | 99 | 146 | 99 |
| Need Security | 174 | 87 | 125 | 84 |
| Social Objectivity | 184 | 92 | 122 | 82 |
| Bell System Value Orientation | 197 | 99 | 142 | 96 |
| Ability to Delay Gratification | 190 | 95 | 142 | 96 |
| Range of Interests | 193 | 97 | 146 | 99 |
| Mean | 179 | 90 | 117 | 79 |

interests. Variables that proved clearly less ratable from the interview reports included personal impact-forcefulness, oral communications skills, and personal impact-likability.

It may be surprising that such variables as personal impact and oral communications skills were not consistently ratable since they are qualities eminently observable in the interview situation. This paradox is resolved by remembering that these variables were freely observable throughout much of the 3½ days of assessment, and the interviewers felt no strong need to report on these characteristics. It must be emphasized that the research reported in this article is not based on an interview designed to cover all variables equally but an interview which was intended to supplement the rest of the assessment process.

The estimated reliabilities of the ratings are presented in Table 2 (Ns for each vari-

able being the same as in Table 1). For the college graduate sample the range is from .73 to .92, with a median of .82. The reliabilities for the noncollege sample tend to be lower, ranging from 0 to .92 with a median of .72. Whereas only 6 of the reliabilities are below .80 in the college graduate sample, 11 fall below .80 in the noncollege sample. The interview reports, it can be seen, yielded not only less information but also less reliable information for the noncollege as compared to the college sample. (Simple practice in coding is not the reason for these results; the college sample was coded first.) Comparison of Tables 1 and 2 indicates that the most complete and reliable information for both samples was obtained on the variables of need advancement, tolerance of uncertainty, social objectivity, and range of interests.

To discover possible clues for the discrepancies between the reliabilities of several of the variables, comparisons were made between the variances of the variables (Table 3; *N*s same as in Table 1). Those *F* ratios with larger variance in the numerator were computed. Statistically reliable differences in the variances of eight of the variables were ob-

TABLE 3
VARIANCES AND *F* RATIOS

| Variable | College sample | Noncollege sample | <i>F</i> |
|--------------------------------|-----------------------|-----------------------|----------|
| | <i>s</i> ² | <i>s</i> ² | |
| Personal Impact–Forcefulness | 3.99 | 4.75 | 1.19 |
| Oral Communication Skills | 4.50 | 5.57 | 1.24 |
| Human Relations Skills | 2.31 | 3.65 | 1.58* |
| Personal Impact–Likability | 2.02 | 2.87 | 1.42 |
| Behavior Flexibility | 2.00 | 1.08 | 1.85** |
| Need Approval–Superiors | 2.10 | 1.97 | 1.07 |
| Need Approval–Peers | 2.71 | 2.04 | 1.33 |
| Tolerance of Uncertainty | 2.70 | 2.96 | 1.10 |
| Inner Work Standards | 2.25 | 1.76 | 1.28 |
| Primacy of Work | 1.67 | 1.74 | 1.04 |
| Energy | 1.73 | 2.38 | 1.38* |
| Goal Flexibility | 3.06 | 2.21 | 1.38* |
| Need Advancement | 2.73 | 3.74 | 1.37* |
| Need Security | 4.28 | 3.25 | 1.32* |
| Social Objectivity | 4.51 | 4.35 | 1.04 |
| Bell System Value Orientation | 1.84 | 1.39 | 1.32* |
| Ability to Delay Gratification | 1.94 | 1.46 | 1.33* |
| Range of Interests | 1.93 | 1.74 | 1.11 |

* .02 < *p* < .10 that variances are equal.
** *p* < .02 that variances are equal.

TABLE 2
CODER RELIABILITIES

| Variable | College sample | Noncollege sample |
|--------------------------------|------------------------|-------------------|
| | <i>r</i> _{ii} | |
| Personal Impact–Forcefulness | .90 | .92 |
| Oral Communication Skills | .92 | .92 |
| Human Relations Skills | .82 | .82 |
| Personal Impact–Likability | .85 | .89 |
| Behavior Flexibility | .77 | .08 |
| Need Approval–Superiors | .79 | .74 |
| Need Approval–Peers | .82 | .67 |
| Tolerance of Uncertainty | .80 | .80 |
| Inner Works Standards | .82 | .67 |
| Primacy of Work | .76 | .72 |
| Energy | .84 | .70 |
| Goal Flexibility | .84 | .55 |
| Need Advancement | .86 | .90 |
| Need Security | .90 | .68 |
| Social Objectivity | .90 | .89 |
| Bell System Value Orientation | .76 | .72 |
| Ability to Delay Gratification | .73 | .00 |
| Range of Interests | .78 | .79 |

tained. In five of these instances the variance for the college sample was larger. In the instance of behavior flexibility, the variance for the college graduate sample is nearly double that for the noncollege sample. The relatively low variance for the latter probably contributes to the low reliability of the ratings (*r*_{ii} = .08) and suggests that for the noncollege sample the coders had difficulty discriminating on this variable. For the remaining variables, however, the possible influence of differences in the variances on discrepancies in the reliabilities of the ratings would be difficult to assess.

The intercorrelations between the interview variables, disregarding signs, range from 0 to .71 (Table 4). Once again the coefficients for the college sample are generally higher than for noncollege men. The median intercorrelation for the former group is .24 as compared to .17 for the latter. The size of these correlations indicates that the interview

TABLE 4
INTERCORRELATIONS OF INTERVIEW VARIABLES

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|-----|----|----|
| Personal Impact-Forcefulness (1) | | | | | | | | | | | | | | | | | | |
| Oral Communications Skills (2) | 19 | | | | | | | | | | | | | | | | | |
| Human Relations Skills (3) | 55 | 30 | | | | | | | | | | | | | | | | |
| Personal Impact-Likability (4) | 09 | 30 | 42 | | | | | | | | | | | | | | | |
| Behavior Flexibility (5) | -07 | -05 | 15 | 24 | | | | | | | | | | | | | | |
| Need Approval-Superiors (6) | -11 | -26 | -45 | -24 | -03 | | | | | | | | | | | | | |
| Need Approval-Peers (7) | -21 | -10 | -20 | 06 | -04 | 51 | | | | | | | | | | | | |
| Tolerance of Uncertainty (8) | 36 | 29 | 71 | 16 | 14 | -58 | -44 | | | | | | | | | | | |
| Inner Work Standards (9) | 28 | 11 | 09 | 14 | 02 | -02 | -01 | 19 | | | | | | | | | | |
| Primacy of Work (10) | 54 | 00 | 36 | 23 | 10 | -23 | -04 | 31 | 43 | | | | | | | | | |
| Energy (11) | 35 | -01 | 36 | 10 | 11 | -27 | -12 | 41 | 34 | 50 | | | | | | | | |
| Goal Flexibility (12) | -40 | -15 | -32 | -04 | 11 | 05 | -10 | -19 | -17 | -28 | -18 | | | | | | | |
| Need Advancement (13) | 63 | 34 | 29 | 04 | 03 | -11 | 03 | 27 | 20 | 48 | 28 | -52 | | | | | | |
| Need Security (14) | -31 | -33 | -04 | 02 | 14 | 13 | 27 | -21 | 07 | -01 | -06 | 19 | -34 | | | | | |
| Social Objectivity (15) | -38 | 21 | 16 | 12 | 26 | -27 | -22 | 11 | 16 | 14 | 24 | 07 | 01 | 02 | | | | |
| Bell System Value Orientation (16) | -11 | -08 | 11 | 05 | 11 | 12 | 21 | -08 | 24 | 26 | 13 | 12 | -04 | 41 | 23 | | | |
| Ability to Delay Gratification (17) | -05 | 13 | 21 | -12 | 06 | -27 | -23 | 38 | 05 | 16 | -05 | -05 | -08 | 02 | 14 | 15 | | |
| Range of Interests (18) | 34 | 31 | 39 | 28 | 23 | -26 | -02 | 36 | 09 | 25 | 39 | -11 | 20 | -22 | 14 | -06 | 13 | |

NOTE.—College sample is above diagonal (N's vary from 85 to 195; median = 166). Noncollege sample is below diagonal (N's vary from 17 to 144; median = 103).

TABLE 5
CORRELATIONS WITH STAFF JUDGMENTS

| Variable | Factor | | | | | | | | | | |
|--------------------------------|--------|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|
| | I | II | III | IV | V | VI | VII | VIII | IX | X | XI |
| College sample | | | | | | | | | | | |
| Personal Impact-Forcefulness | 52 | 52 | -32 | 42 | 44 | 29 | 22 | 36 | -38 | -16 | 22 |
| Oral Communication Skills | 46 | 46 | -23 | 33 | 36 | 23 | 40 | 26 | -26 | -06 | 32 |
| Human Relations Skills | 29 | 27 | -20 | 16 | 28 | 23 | 06 | 19 | -25 | -14 | 14 |
| Personal Impact-Likability | 27 | 33 | -00 | 22 | 39 | 13 | 04 | 31 | -03 | -01 | 08 |
| Behavior Flexibility | 23 | 22 | -18 | 12 | 21 | 18 | 09 | 10 | -16 | -12 | 17 |
| Need Approval-Superiors | -17 | -09 | 27 | -11 | -05 | -14 | -08 | 03 | 20 | 27 | -24 |
| Need Approval-Peers | -27 | -21 | 30 | -22 | -12 | -16 | -16 | -08 | 24 | 25 | -21 |
| Tolerance of Uncertainty | 31 | 26 | -27 | 17 | 20 | 30 | 18 | 15 | -26 | -20 | 26 |
| Inner Work Standards | 24 | 26 | -08 | 10 | 18 | 19 | 09 | 40 | -16 | -03 | 03 |
| Primacy of Work | 25 | 25 | -16 | 13 | 16 | 19 | 05 | 35 | -22 | -10 | 03 |
| Energy | 32 | 30 | -25 | 20 | 30 | 26 | 05 | 24 | -32 | -15 | 14 |
| Goal Flexibility | -30 | -27 | 22 | -20 | -20 | -19 | -13 | -31 | 24 | 13 | -02 |
| Need Advancement | 41 | 30 | -43 | 27 | 23 | 23 | 23 | 27 | -57 | -24 | 33 |
| Need Security | -40 | -30 | 44 | -17 | -23 | -23 | -29 | -19 | 50 | 21 | -47 |
| Social Objectivity | 16 | 19 | 00 | 10 | 10 | 13 | 19 | 17 | 00 | -01 | 28 |
| Bell System Value Orientation | -11 | -07 | 13 | -06 | -03 | -06 | -17 | 01 | 12 | 05 | -18 |
| Ability to Delay Gratification | -04 | -03 | 05 | 00 | -06 | 04 | 02 | -03 | 10 | -01 | 01 |
| Range of Interests | 23 | 27 | 01 | 14 | 24 | 13 | 35 | 15 | -05 | 11 | 17 |
| Noncollege sample | | | | | | | | | | | |
| Personal Impact-Forcefulness | 48 | | | 24 | 25 | 25 | 03 | 50 | -58 | -25 | |
| Oral Communication Skills | 44 | | | 53 | 36 | 24 | 47 | 23 | -24 | -19 | |
| Human Relations Skills | 43 | | | 32 | 46 | 45 | 10 | 14 | -06 | -11 | |
| Personal Impact-Likability | 06 | | | 07 | 09 | 09 | 13 | -02 | 11 | -09 | |
| Behavior Flexibility | 06 | | | -02 | 15 | 00 | 06 | -01 | 06 | 14 | |
| Need Approval-Superiors | -24 | | | -20 | -23 | -34 | -22 | -10 | -03 | 07 | |
| Need Approval-Peers | -23 | | | -25 | -17 | -22 | -29 | -15 | 02 | 00 | |
| Tolerance of Uncertainty | 29 | | | 21 | 31 | 40 | 18 | 20 | -08 | -23 | |
| Inner Work Standards | 09 | | | 01 | 07 | 01 | -03 | 43 | -05 | 17 | |
| Primacy of Work | 21 | | | 03 | 20 | 13 | -19 | 32 | -25 | 10 | |
| Energy | 32 | | | 05 | 31 | 21 | 05 | 35 | -15 | 00 | |
| Goal Flexibility | -20 | | | -16 | -14 | -16 | 10 | -17 | 41 | 17 | |
| Need Advancement | 40 | | | 29 | 27 | 19 | -14 | 31 | -67 | -06 | |
| Need Security | -25 | | | -29 | -10 | -05 | -22 | -03 | 37 | 12 | |
| Social Objectivity | 19 | | | 18 | 19 | 15 | 19 | 13 | 02 | 28 | |
| Bell System Value Orientation | -07 | | | -16 | 02 | 03 | -13 | 06 | 20 | 18 | |
| Ability to Delay Gratification | 02 | | | 10 | 03 | 26 | 11 | 04 | 15 | -12 | |
| Range of Interests | 19 | | | 19 | 19 | 14 | 33 | 11 | -10 | -04 | |

variables were relatively independent. Nevertheless, some clustering of the variables, particularly in the college graduate sample, seems apparent. One cluster (personal impact-forcefulness, oral communications skills, and human relations skills) incorporates variables reflecting interpersonal skills. Another (needs for approval-superiors and peers) reflects dependence on others. A third (inner work

standards, primacy of work and energy) concerns work motivation. Finally, the negative correlation between need advancement and need security suggests career motivation. Of the remaining variables several (i.e., personal impact-likability, behavior flexibility, tolerance of uncertainty, and goal flexibility) have fairly substantial correlations with variables in one or more of the clusters, which in turn

end to overlap considerably with each other. Four variables (social objectivity, Bell System value orientation, ability to delay gratification, and range of interests) have relatively low correlations with all other variables.

The main question to which this article is addressed is the role of the interview in the total assessment center process. The correlations between each of the interview variables and variables reflecting assessment staff judgments made on the basis of all the assessment techniques are shown in Table 5 (*Ns* same as in Table 1).

Scores based on factorial analyses of 25 characteristics rated by the staff (Bray & Grant, 1966) follow. The characteristics, selected for their relevance to the study, include managerial skills, interpersonal relationships, general abilities, motives, values, and attitudes. The factors obtained and their designations are

| Factor | Identification |
|---------------------------|--------------------------|
| I | General Effectiveness |
| II (college sample only) | General Effectiveness |
| III (college sample only) | Passive Dependency |
| IV | Administrative Skills |
| V | Interpersonal Skills |
| VI | Control of Feelings |
| VII | Intellectual Ability |
| VIII | Work-oriented Motivation |
| IX | Passivity |
| X | Dependency |
| XI (college sample only) | Nonconformity |

Table 5 shows that judgments of personal characteristics based on the interview reports alone and made independently of the assessment staff judgments, which were based on information from all of the assessment techniques, correlate substantially with assessment staff judgments. Furthermore, the consistencies in magnitude and direction of the correlations from sample to sample are relatively high.

To assist in interpreting these data, interview variables correlating .30 or higher with a staff judgment factor for *both* the college

and noncollege samples are listed below under each staff judgment factor:

- I. General Effectiveness
 - Personal Impact-Forcefulness
 - Oral Communication Skills
 - Need for Advancement
 - Energy
- IV. Administrative Skills
 - Oral Communication Skills
- V. Interpersonal Skills
 - Oral Communication Skills
 - Energy
- VI. Control of Feelings
 - Tolerance of Uncertainty
- VII. Intellectual Ability
 - Oral Communications Skills
 - Range of Interests
- VIII. Work-oriented Motivation
 - Personal Impact-Forcefulness
 - Inner Work Standards
 - Primacy of Work
- IX. Passivity
 - Need for Advancement (negative)
 - Personal Impact-Forcefulness (negative)
 - Need for Security
- X. Dependency
 - (none)

It is apparent from the above that some of the characteristics judged solely on the basis of interview reports do correlate substantially with staff judgments based on all the assessment techniques. It is also clear that some interview variables showed more pronounced relationships than others. Personal impact-forcefulness, oral communication skills, energy, and need advancement, for example, were obviously potent interview variables.

On the whole, the correlations for the selected variables are meaningful. One might expect, for example, that judgments of administrative skills, interpersonal skills, and intellectual ability would relate to ratings based on interview information for oral communications skills. Most of the remaining relationships make sense. The only correlations which are difficult to interpret are those between personal impact-forcefulness and work-oriented motivation.

Just how much influence the interview may have had on assessment staff judgments can-

not be ascertained from these data. Comparison of the correlations in Table 5 with those for other assessment techniques (Grant, Katkovsky, & Bray, 1967, p. 231), however, shows that the interview appears to have played a primary role in the judgment of the factor of passivity. In addition, the correlations between some of the interview variables and other factors are of the same magnitude as those for other assessment techniques. Since most assessment staff members are convinced that the interview report does not influence their ratings as much as behavior in the simulations or test results, these correlations are probably not produced by the influence of the interviewer report. Instead, the interview gets successfully at some of the same dimensions, though not necessarily influencing judgments.

A staff judgment on which the interview report clearly has a direct influence is that of passivity. This factor, which might be better labeled "career passivity," involves the lack of a strong need to advance in the organization, a willingness to wait for advancement, and an emphasis on job security. The correlation of the interview variable of need advancement with this factor for both

the college and the noncollege groups ($-.57$ and $-.67$) is higher than for any other assessment measure.

A complete evaluation of the relative contributions of the interview to the assessment process awaits further studies. Regression analyses of all assessment techniques with scores based on the factorially derived characteristics are under way. In addition, analyses relating interviews and other assessment technique variables to the 25 characteristics directly rated by the assessment staff are planned. Though scores based on the factorially derived characteristics have proved useful in making studies of the interview and other assessment techniques, information pertinent to relatively specific characteristics (e.g., Bell System Value Orientation) has undoubtedly been omitted in the process.

In addition to rating 25 qualities (those summarized in the factors of Table 5), the assessment staff made predictions of progress in the management hierarchy. The specific prediction was whether each *S* would reach middle management within 10 yr. from the time of assessment. These predictions have proved significantly accurate (Bray & Grant, 1966). Table 6 (same *Ns* as in Table 1) shows the correlation of each interview variable with this staff prediction.

An inspection of the table reveals that 22 of the 36 correlations are statistically significant at the .05 level. Once again the present data cannot reveal the extent to which the staff was influenced by the interview report, but it is clear that the interview successfully captured characteristics relevant to the total assessment process. Significant for both the college and noncollege samples were the interview variables of oral communication skills, human relations skills, tolerance of uncertainty, goal flexibility, need advancement, and range of interests.

Studies of the overlap of the interview with other assessment techniques must also await a total analysis of the assessment process. It may be interesting, however, to note the relationship in the college sample of one of the more important interview variables, personal impact-forcefulness, to some of the other assessment measures. Forcefulness in the interview correlates .35 with contribution to

TABLE 6
CORRELATIONS WITH STAFF PREDICTIONS

| Variable | College sample | Noncollege sample |
|--------------------------------|----------------|-------------------|
| Personnel Impact-Forcefulness | .49* | .21 |
| Oral Communication Skills | .41* | .48* |
| Human Relations Skills | .23* | .38* |
| Personal Impact-Likability | .25* | .14 |
| Behavior Flexibility | .19* | .11 |
| Need Approval-Superiors | -.02 | -.20* |
| Need Approval-Peers | -.21* | -.13 |
| Tolerance of Uncertainty | .23* | .36* |
| Inner Work Standards | .17* | .07 |
| Primacy of Work | .20* | .21 |
| Energy | .25* | .17 |
| Goal Flexibility | -.30* | -.21* |
| Need Advancement | .28* | .42* |
| Need Security | -.28* | -.17 |
| Social Objectivity | .03 | .18* |
| Bell System Value Orientation | -.13 | -.05 |
| Ability to Delay Gratification | .01 | .03 |
| Range of Interests | .27* | .26* |

* $p < .05$ that $p = .00$.

TABLE 7
CORRELATIONS WITH SALARY PROGRESS

| Interview variable | College graduates | | Noncollege | |
|---------------------------|-------------------|-------|------------|-------|
| | N | r | N | r |
| Personal Impact— | | | | |
| Forcefulness | 71 | .28* | 43 | .17 |
| Oral Communication Skills | 69 | .22 | 69 | .50* |
| Human Relations Skills | 76 | .20 | 48 | .41* |
| Personal Impact— | | | | |
| Likability | 50 | .22 | 37 | -.11 |
| Behavior Flexibility | 80 | .30* | 119 | .04 |
| Need Approval—Superiors | 71 | -.36* | 102 | -.27* |
| Need Approval—Peers | 75 | -.36* | 96 | -.17 |
| Tolerance of Uncertainty | 74 | .06 | 105 | .23* |
| Inner Work Standards | 80 | .07 | 114 | .08 |
| Primacy of Work | 75 | .30* | 118 | .25* |
| Energy | 74 | .35* | 100 | .16 |
| Goal Flexibility | 74 | -.08 | 120 | -.16 |
| Need Advancement | 80 | .49* | 120 | .44* |
| Need Security | 75 | -.35* | 112 | -.26* |
| Social Objectivity | 75 | .17 | 101 | .20* |
| Bell System Value | | | | |
| Orientation | 80 | -.11 | 120 | -.10 |
| Ability to Delay | | | | |
| Gratification | 76 | -.16 | 116 | .12 |
| Range of Interests | 79 | .28* | 121 | .22* |

* $p < .05$ that $p = .00$.

the Manufacturing Problem, .49 with contribution to the Group Discussion Problem, .35 with Projective Test achievement motivation, .35 with Projective Test willingness to assume a leadership role, -.30 with Projective Test dependence, .32 with need dominance on the Edwards Personal Preference Inventory, and .32 with ascendance on the Guilford-Martin Inventory. (These are the highest seven correlations of this interview variable with 41 measures from other techniques for the college sample.)

A final question is the extent to which the interview report ratings are directly related to progress in management. The data in Table 7 concern 81 college graduates having 8–10 yr. of experience in two telephone companies and on 122 noncollege men with 8–9 yr. of experience in two such companies. The correlations shown are average correlations for the two company samples. The progress criterion was obtained by computing the difference between each S's salary at the

time of assessment and his salary on June 30, 1967.

Eighteen of the 36 coefficients in this table are significant at the .05 level, 9 each for the college and noncollege groups. Interview variables reliably predictive for both groups were need approval—superiors (negative), primacy of work, need advancement, need security (negative), and range of interests.

These results are comparable to those relating the interview report variables to assessment staff evaluations, i.e., the staff predictions (Table 6) and scores based on the general effectiveness factor (Table 5). Though the specific correlations vary, the patterns of correlations are roughly similar. The findings thus indicate that the assessment staff was interpreting organizational values correctly.

DISCUSSION

The results of this investigation of the assessment interview clearly indicate that the interview reports contributed to the assessment process. Judgments of career motivation apparently depended heavily on the interview. Work motivation and control of feelings ratings also appear to have been influenced by the interview information. In addition, judgments of interpersonal skills were at least reinforced, if not influenced, by the interview reports.

That information from the interview was also predictive is demonstrated by the large number of statistically significant correlations with success in management. Variables reflecting career motivation, dependency needs, work motivation, and interpersonal skills were related to individual differences in salary increases.

As noted previously, this investigation does not in itself establish the relative weight of the interview as compared to the other techniques in the assessment center process. It is, furthermore, not an experimental test of how much the interview could accomplish. The interviewers made, for example, no efforts to uncover information on administrative skills since this was presumably adequately covered elsewhere. The results demonstrate, nevertheless, that the interview did produce reliable ratings of managerial qualities which cor-

related significantly with ratings made on the basis of several other techniques and with advancement.

The method used in quantifying the interview data is, of course, more characteristic of that used in analyzing interview information obtained from surveys than that used in studying assessment or selection interviews. The ratings of selected variables were made by independent coders rather than by the persons doing the interviews. Whether interviewer ratings would have produced similar results cannot be determined. The findings are pertinent to a suggestion by Ulrich and Trumbo (1965), however, that the selection interviewer should function as an information gatherer and reporter, leaving selection decisions up to others.

With regard to various issues in interviewing raised by reviewers of research on the topic (Mayfield, 1964; Ulrich & Trumbo, 1965), the findings of this study tend to shed either light or confusion, depending on one's point of view. In contrast to their findings, the interview reports from relatively unstructured interviews yielded quite reliable and valid (i.e., predictive) information. Furthermore, the information on many of the variables was sufficiently complete for two coders to make ratings. Perhaps the issue is not that of structured vs. unstructured interviews but of interviewer skill and understanding of what is to be covered.

Finally, with regard to the personal characteristics an interviewer can identify, the findings of this study tend to be more supportive of the views of Ulrich and Trumbo (1965) than of Mayfield (1964). The interviewers did identify career motivation and interpersonal skill characteristics. They apparently also obtained reliable information on additional characteristics. They were not asked to obtain information on intellectual abilities per se, though some of the characteristics identified did correlate substantially with

assessment staff judgments of intellectual abilities and with mental ability test scores.

The findings of this study give positive support to the use of the interview in assessment center procedures. They also suggest several possible areas for research on the interview. Among such would be studies comparing judgments of interviewers themselves with those made by others on the basis of the interview reports. Such studies would bear on the issue of whether interviewers should be primarily reporters or assessors. Additional studies on the personal characteristics which an interviewer can identify would also be useful. The findings of this study suggest that the interview may have considerable scope and still be reliable and valid.

REFERENCES

- BRAY, D. W. The management progress study. *American Psychologist*, 1964, **19**, 419-420.
- BRAY, D. W., & GRANT, D. L. The assessment center in the measurement of potential for business management. *Psychological Monographs*, 1966, **80** (17, Whole No. 625).
- GRANT, D. L., KATKOVSKY, W., & BRAY, D. W. Contributions of projective techniques to assessment of management potential. *Journal of Applied Psychology*, 1967, **51**, 226-232.
- KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: University of Michigan, 1951.
- MAYFIELD, E. C. The selection interview—a re-evaluation of published research. *Personnel Psychology*, 1964, **17**, 239-260.
- MACKINNON, D. W. An assessment study of Air Force officers. Part V: Summary and applications. Technical Report WADC-TR-59-91 (V), ASTIA Document No. AD 210 220, 1958. Office of Technical Services, United States Department of Commerce.
- Office of Strategic Services in World War II. Assessment Staff. *Assessment of men*. New York: Rinehart, 1948.
- PRIEN, E. P. Assessments of higher-level personnel. V: An analysis of interviewers' predictions of job performance. *Personnel Psychology*, 1962, **15**, 319-334.
- ULRICH, L., & TRUMBO, D. The selection interview since 1949. *Psychological Bulletin*, 1965, **63**, 100-116.

(Early publication received August 21, 1968)

CONSISTENCY AND GENERALIZABILITY OF INTRAINDIVIDUAL VARIABILITY¹

RALPH F. BERDIE²

Office of the Dean of Students, University of Minnesota

In order to observe the extent to which a person's behavior is variable on 1 set of tasks and on another set of similar tasks, and to observe the extent to which his behavior is variable on 1 set of tasks and on a set of dissimilar tasks, 100 Ss were given 6 tests, a different form of each test being administered on 20 successive days. Variance indices, reflecting intraindividual variability over time, were derived from the even- and odd-numbered test forms, and the correlations between variability on the even- and odd-numbered forms ranged from .25 to .89. Variance indices on the 6 different tests correlated with one another from -.23 to .53. Temporal intraindividual variability on some tasks can be reliably and meaningfully observed. Such variability is not widely generalized over a large portion of the behavior domain and neither is it specific to each task or behavior. Variability on some tasks is related to variability on certain other tasks.

The extent to which an individual's behavior is consistently variable on a given task or on different tasks may provide cues as to the effectiveness of his behavior. The variability studied here corresponds to that described by Fiske and Maddi (1961): "one form of variability, the variation in the behavior of a given organism at different times but under the same external conditions [p. 327]." This form of variability is to be contrasted with that discussed by Hull (1927) who was concerned with variability in the amount of different traits possessed by an individual, and also to be contrasted with the variability discussed by Wechsler (1952) who was concerned with the variability of a given trait within the population.

Theoretically, an individual cannot reproduce a behavior identically, since once he has performed a task, repetition of that task must be influenced by its prior performance. In spite of the impossibility of studying the individual's variability while performing the same task, his variability can be studied while performing a homogeneous group of tasks.

Tasks and situations change from performance to performance, but these can be ordered into highly similar categories and variability of behavior so studied.

Fiske provided one theoretical basis for the analysis of variability. He regarded individual variability as having a coping function which affects the organism's adaptability. When several alternate behaviors are available to the organism, its inherent variability increases the likelihood that the organism eventually will select and adopt the response which best copes with a given situation. Fiske and Maddi (1961) defined several questions concerning individual differences in variability.

In an earlier study, Berdie (1961) found that variability on a task involving advanced high school mathematics appeared consistent within the individual and that this variability might be related to the extent to which a person's college achievement could be predicted on the basis of aptitude tests. The two questions approached in the present study were: (1) Can the variability over time of an individual's behavior be reliably and consistently observed? (2) To what extent are persons variable over time on one task also variable on other tasks?

If variability over time can be reliably observed and is not specific to each task, then the variability of an individual may be a useful concept in understanding his behavior. For example, variable persons may be more

¹ This project was supported with funds from the United States Office of Education, Office of Educational Grant 3-7-068694-2082. Appreciation is expressed to Richard Arvey and Diane Johnson Tinsley for their contribution in analyzing the data.

² Requests for reprints should be sent to the author, Student Life Studies, Office of the Dean of Students, University of Minnesota, 2001 Riverside Avenue, Minneapolis, Minnesota 55455.

or less predictable or more or less task oriented.

METHOD

Tasks

The behavior observed consisted of responses to six of the Repetitive Psychometric Measures (scores on these tests) developed by Moran and Mefferd (1959). Each of the six tests represents a distinct factor derived from repeated factor-analytic studies. The Aiming test consists of 15 rows each containing 20 circles, and the circles are in rows connected sequentially by a line. The *S* places the test on a piece of corrugated paper and the task is to punch a hole inside as many circles as possible within 90 sec. without touching the circles. The *Ss* in this experiment used a stylus consisting of a pencil-sized piece of wood with a thin pinpoint at one end. The test involves the ability to carry out quickly and precisely a series of movements depending on eye-hand coordination.

The Flexibility of Closure test requires *S* to copy 36 geometric figures into matrices of dots. Each test form contains 36 figures. The task, as described by the authors, is to retain the image of a specified configuration despite the influence of other distracting configurations in the perceptual field. The Number Facility test is similar to French's *N* factor and consists of 90 problems each requiring the addition of three two-digit numbers.

The Perceptual Speed test requires the identification of well-known symbols in a mass of material and consists of rows of 30 digits with an encircled digit at the left of each row. The task is to cross out every digit in the row similar to the encircled digit. The time limit specified by Moran and Mefferd is 2½ min., but early experience with this test suggested that too many *Ss* completed the test within this time limit and the time limit was reduced to 1½ min.

The Speed of Closure test measures the ability to unify an apparently disparate perceptual field into a single percept. Each form consists of 22 lines and each line has letters in it apparently arranged at random but containing from two to four 4-letter words which are to be encircled. The final test, Visualization, consists of tangled lines which must be followed visually from their start to finish.

For each of these tests Moran and Mefferd developed 20 different forms with the original intent that the forms would be equivalent. Later study, however, indicated that on each of the tests but Number Facility the alternate forms were reliably different (Moran, Kimble, & Mefferd, 1964) and correction factors were provided for the 20 alternate forms of these five tests. These correction factors were not used in this study, in light of the experimental design.

The test authors, comparing scores on Form 1 and Form 2, reported test-retest reliabilities ranging from .72 to .94. Intercorrelations of the six tests, using only Form 1, ranged from .09 to .44. Considering the purpose for which these tests are to be used here,

they appeared to be adequately reliable and sufficiently independent from one another.

Sample

The population from which the *Ss* were drawn consisted of freshmen entering the University of Minnesota Institute of Technology in the fall of 1966. All freshmen were informed of the possibility of participating in the experiment and, from those who volunteered, *Ss* were selected on the basis of class schedules, availability of data, and proximity to campus. The *Ss* consisted of fairly representative bright college students who had survived at least one demanding academic quarter and who were motivated to earn \$40 by participating in an experiment that would cause them no stress or discomfort.

Procedures

The experiment was conducted in a well-isolated subbasement room with overhead lights and lamps arranged so that illumination was not brilliant but *Ss* could see comfortably. No noise from outside the building penetrated the room and little traffic passed in the corridor outside of the door. Temperature in the room was constant and comfortable, although when the door was closed there was little ventilation. Insofar as *Ss* remained in the room for periods of only 20 min. and the door was kept open for at least ½ hr. between sessions, lack of ventilation produced no discomfort.

The *Ss* were seated in the center of the room in classroom chairs with arm tablets. They were divided into five groups and a group was tested each day at 9:30 A.M., 12:30 noon, 1:30, 2:30, and 3:30 P.M. Assignments to time periods were based on the class schedules submitted by *Ss*.

Approximately one-fifth of the *Ss* took Form 1 of the test on the first day, Form 2 on the second, etc. Another group of *Ss* took Form 5 on the first day, Form 6 on the second day, and on the twentieth day took Form 4. Other groups of *Ss* started on Form 9, Form 13, and Form 17, in order to provide some randomization of form-sequence influence. Within each time session, students were randomly assigned to sequence groups.

At the first session, the experimenter read to each group an introductory statement, and a trained and experienced psychometrist then read the test instructions, administered the practice exercises provided by Moran and Mefferd, and administered the tests.

Testing schedules for each group were arranged Monday through Friday for 4 successive wk., and *Ss* who missed sessions made them up during an adjacent session or during the fifth week. Of the 100 *Ss*, 62 attended daily. Over 95% of the tests were administered to *Ss* at the time of day originally scheduled.

At the completion of the last form of the last test each *S* completed a questionnaire reporting his reaction to the tasks and his perceptions of the purpose of the experiment. The *Ss* were told at the first

session that the purpose of the experiment was to compare the psychological characteristics, as measured by these tests, of students in technology and science to those of other students.

Analysis

The 12,000 test papers were scored by research assistants and, when scoring was completed, 200 papers for each test representing all of the 20 forms were drawn and rescored. The original scores were compared to those obtained by rescoring, and frequencies of errors of various sizes were tabulated and correlations determined. Scoring was judged to be adequate for five of the six tests but all 2,000 of the Number Facility tests were rescored. The test scores were then entered on basic record cards for each student, verified, and then punched and verified on IBM cards.

For each of the six tests a variability index was computed for each student. This consisted of the variance (SD^2) of the 20 raw scores derived from the 20 forms of the tests. At the same time, for each of the six tests a mean score was computed for each student, this consisting of the mean of the 20 scores derived from the 20 forms.

Then, in order to facilitate comparisons between tests and to provide a basis for obtaining a total variance index, each raw score was transformed to a standard score, using a mean equivalent to 50 and a standard deviation equivalent to 10, based on the distribution of 100 scores of each form of each test. For example, the 100 scores on Form 1 of the Aiming test were selected, the mean and standard deviation calculated, and, for each student, his raw score on Form 1 was transformed to a standard score based on this distribution. Then, for each of the six tests a variability index for each student was computed, along with a mean index, based on the 20 standard scores. A seventh variance and mean index were calculated for each student, based on all 120 of the scores.

Thus, for each student six variance indices and six mean indices based on raw scores were available, and seven variance indices and seven mean indices were available based on standard scores.

The consistency of the variability index was revealed by what corresponds to an odd/even reliability coefficient. The scores on each of the 10 odd-numbered forms were used to provide a variance and a mean, and the scores on each of the 10 even-numbered forms provided comparable indices. On each test, each *S* had two variance indices and two mean indices, and in each instance the correlation was calculated between the even-numbered-form and the odd-numbered-form indices. The analysis was done first using raw scores and then standard scores.

The correlation coefficients then were calculated between the variance indices on the six tests, first using the indices based on raw scores and then the indices based on standard scores. The mean indices were analyzed similarly.

RESULTS

Reliability of Indices

Table 1 shows the correlations between variance and mean indices based on odd-numbered and even-numbered forms, using raw scores. Table 2 presents similar information based on standard scores and includes information on the total score, which consists of the sum of the standard scores for the six tests. The test scores are reliable, as shown by the mean score correlations which range from .96 to .99.

Two of the variance indices, one based on Aiming and the other on Number Facility, show relatively high consistency; two, Perceptual Speed and Speed of Closure, provide correlation coefficients in the mid-50s. Visualization provides the lowest reliability coefficient, .25. The total variance index provides a correlation of .89, suggesting that whatever this is it is an index that can be obtained rather consistently.

These estimates of reliability are based on 10 scores. When one uses the Spearman-Brown prophecy formula, reliability estimates of the variability index based on 20 scores are: Aiming, .91; Flexibility of Closure, .58; Number Facility, .89; Perceptual Speed, .71; Speed of Closure, .73; and Visualization, .40. One can conclude that, on Aiming and Number Facility, variability within persons tends to be remarkably consistent and consistency of variability is found on all other tests.

Berdie (1961), in a previous study, observed the reliability of a similar variance measure based on 10 subscores of a mathematics achievement test. The reliability coefficients for various groups in that study ranged about .90. Using a somewhat similar method of analyses of varied repeated personality assessment data, Fiske (1957) reported odd/even reliabilities extending from .46 to .96. His results also suggested that the extent of consistency of a person's variability depended in part on the task or instrument used.

Relationships between Variance Indices

Table 3 shows the intercorrelations between the six variance indices based on raw scores and Table 4 similar intercorrelations of the seven indices, including the total variance in-

TABLE 1

CORRELATIONS FOR 100 UNIVERSITY OF MINNESOTA INSTITUTE OF TECHNOLOGY FRESHMEN ON VARIANCE INDEX AND MEAN RAW SCORE BASED ON 10 ODD-NUMBERED FORMS AND 10 EVEN-NUMBERED FORMS OF EACH OF SIX REPETITIVE PSYCHOMETRIC MEASURES TESTS

| Test | Variance index | | | | | Mean raw score | | | | |
|------------------------|----------------|--------------|-----------|---------------|-----------|----------------|--------------|-----------|---------------|-----------|
| | <i>r</i> | Odd-numbered | | Even-numbered | | <i>r</i> | Odd-numbered | | Even-numbered | |
| | | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Aiming | .83 | 210.74 | 182.04 | 199.94 | 186.46 | .96 | 111.40 | 15.64 | 114.34 | 16.13 |
| Flexibility of Closure | .41 | 20.16 | 11.79 | 17.55 | 10.57 | .97 | 18.29 | 4.77 | 20.10 | 5.16 |
| Number Facility | .80 | 29.54 | 29.45 | 24.44 | 26.11 | .99 | 43.04 | 10.34 | 44.46 | 10.59 |
| Perceptual Speed | .55 | 33.28 | 20.63 | 44.12 | 30.67 | .97 | 61.64 | 5.53 | 63.84 | 6.18 |
| Speed of Closure | .57 | 69.36 | 33.27 | 52.85 | 28.14 | .98 | 38.85 | 6.43 | 39.31 | 6.38 |
| Visualization | .25 | 47.67 | 25.27 | 38.31 | 23.55 | .98 | 53.84 | 7.80 | 54.28 | 8.02 |

Note.—All correlations significant beyond .01 level except one significant between .01 and .05.

dex, based on standard scores. In Table 3, of the 15 correlations, 4 were significant beyond the .01 level of probability, 1 between the .05 and .01 level. The variances for Aiming and Number Facility correlated .47, between Aiming and Speed of Closure, .28, between Speed of Closure and Number Facility, .22, and between Number Facility and Visualization, .27. The highest intercorrelation was found between the two variance indices having the highest reliability, and the intercorrelations must be examined in light of the reliabilities of the variance indices.

Using the uncorrected reliabilities based on the odd- and even-numbered forms and cor-

recting the intertest variance correlations for attenuation (unreliability), the correlation between the variances for Aiming and Number Facility increases from .47 to .58, between Aiming and Speed of Closure from .28 to .41, between Number Facility and Speed of Closure from .22 to .33, between Number Facility and Visualization from .27 to .60. These correlations suggest that some of the observed independence between the variances is due to the unreliabilities of the variance indices.

The reliability coefficients themselves are minimum estimates and one can correct them as we have done before by ap-

TABLE 2

CORRELATIONS FOR 100 UNIVERSITY OF MINNESOTA INSTITUTE OF TECHNOLOGY FRESHMEN OF VARIANCE INDEX AND MEAN *T* SCORE BASED ON 10 ODD-NUMBERED AND 10 EVEN-NUMBERED FORMS FOR SIX REPETITIVE PSYCHOMETRIC MEASURES TESTS AND FOR "TOTAL MEASURE"

| Test | Variance index | | | | | Mean <i>T</i> score | | | | |
|------------------------|----------------|--------------|-----------|---------------|-----------|---------------------|--------------|-----------|---------------|-----------|
| | <i>r</i> | Odd-numbered | | Even-numbered | | <i>r</i> | Odd-numbered | | Even-numbered | |
| | | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Aiming | .82 | 47.60 | 40.04 | 44.65 | 41.12 | .96 | 49.99 | 7.53 | 50.00 | 7.71 |
| Flexibility of Closure | .42 | 47.03 | 27.46 | 37.28 | 19.04 | .97 | 50.00 | 7.56 | 50.00 | 8.13 |
| Number Facility | .80 | 20.68 | 21.59 | 17.05 | 19.19 | .99 | 50.00 | 9.01 | 50.00 | 9.19 |
| Perceptual Speed | .53 | 42.84 | 33.91 | 37.42 | 32.75 | .97 | 50.00 | 7.81 | 50.00 | 8.12 |
| Speed of Closure | .71 | 61.42 | 30.35 | 52.33 | 27.04 | .98 | 50.00 | 6.65 | 50.00 | 7.24 |
| Visualization | .37 | 39.26 | 22.26 | 34.38 | 21.20 | .98 | 50.00 | 8.02 | 50.00 | 8.29 |
| Total Measure | .89 | 66.91 | 26.56 | 63.15 | 27.37 | .99 | 50.00 | 5.79 | 50.00 | 6.10 |

Note.—All correlations significant beyond .01 level.

TABLE 3

INTERCORRELATIONS BETWEEN THE SIX "VARIANCE" INDICES (USING RAW SCORES) FOR 100 UNIVERSITY OF MINNESOTA INSTITUTE OF TECHNOLOGY FRESHMEN EACH TAKING 20 FORMS OF EACH OF SIX REPETITIVE PSYCHOMETRIC MEASURES TESTS

| Test | FC | NF | PS | SC | V | <i>M</i> | <i>SD</i> |
|-----------------------------|------|-------|------|-------|-------|----------|-----------|
| Aiming (A) | -.02 | .47** | .01 | .28** | .01 | 202.04 | 167.48 |
| Flexibility of Closure (FC) | | .17 | -.15 | .16 | .32** | 19.09 | 9.45 |
| Number Facility (NF) | | | -.18 | .22* | .27** | 26.58 | 25.11 |
| Perceptual Speed (PS) | | | | .14 | -.01 | 38.11 | 21.34 |
| Speed of Closure (SC) | | | | | .06 | 58.45 | 25.88 |
| Visualization (V) | | | | | | 41.59 | 18.23 |

* $p < .05$.

** $p < .01$.

plying the Spearman-Brown prophecy formula. Unlimited corrections of this sort to statistical data lead to a morass of difficulty, particularly when one is concerned with prediction, but in this instance the concern is with arriving at some estimate as to relationships between variances, and these should be based on the best reliability estimates. Table 5 shows the intercorrelations of variance indices when the correlations are corrected for attenuation and the reliability coefficients used have been corrected with the Spearman-Brown prophecy formula. This table gives an optimal estimate of the relationships.

Recognizing the questionable assumptions that have to be made with these two corrections entering into the coefficients, the table reveals that the variance on each of the tests is to some extent related to the variance on one or more of the other tests. The variance index on Number Facility is significantly cor-

related with the index of each of the other five tests. The variances on Flexibility of Closure and Speed of Closure are related to variances on four of the other five tests. Perceptual Speed variance is correlated to three of the other indices, and two of the coefficients are negative, and Aiming and Visualization variances each are correlated with two of the other variance indices. The Aiming and Number Facility indices have the highest reliabilities, are the most highly intercorrelated, and the Number Facility index is significantly related with each other index, although the correlations are small. The best indication of the variance domain may be provided by the Number Facility and Aiming tests.

DISCUSSION

These results suggest that intraindividual variability is not specific to each task and neither does a strongly generalized character-

TABLE 4

INTERCORRELATIONS BETWEEN THE SEVEN "VARIANCE" INDICES (USING *T* SCORES) FOR 100 UNIVERSITY OF MINNESOTA INSTITUTE OF TECHNOLOGY FRESHMEN EACH TAKING 20 FORMS OF EACH OF SIX REPETITIVE PSYCHOMETRIC MEASURES TESTS (ALSO INCLUDED IS THE "TOTAL VARIANCE" INDEX)

| Test | FC | NF | PS | SC | V | TV | <i>M</i> | <i>SD</i> |
|-----------------------------|-----|-------|------|-------|-------|-------|----------|-----------|
| Aiming (A) | .00 | .49** | -.06 | .26** | .05 | .42** | 44.89 | 36.62 |
| Flexibility of Closure (FC) | | .14 | -.10 | .22* | .27** | .15 | 40.84 | 19.25 |
| Number Facility (NF) | | | -.18 | .20* | .32** | .50** | 18.23 | 18.38 |
| Perceptual Speed (PS) | | | | -.01 | .03 | -.03 | 39.15 | 28.02 |
| Speed of Closure (SC) | | | | | .03 | .21* | 54.60 | 25.26 |
| Visualization (V) | | | | | | .28** | 35.76 | 16.97 |
| Total Variance (TV) | | | | | | | 64.69 | 25.97 |

* $p < .05$.

** $p < .01$.

TABLE 5
INTERCORRELATIONS BETWEEN THE SIX VARIANCE INDICES USING RAW SCORES: CORRELATIONS CORRECTED FOR ATTENUATION (UNRELIABILITY OF VARIANCE INDICES) USING SPEARMAN-BROWN CORRECTED RELIABILITY COEFFICIENTS

| Test | FC | NF | PS | SC | V |
|-----------------------------|------|-----|------|-----|------|
| Aiming (A) | -.03 | .53 | .01 | .35 | .02 |
| Flexibility of Closure (FC) | | .24 | -.23 | .25 | .67 |
| Number Facility (NF) | | | -.23 | .28 | .46 |
| Perceptual Speed (PS) | | | | .20 | -.02 |
| Speed of Closure (SC) | | | | | .11 |

Note.—N = 100.

istic of variability extend over a broad variety of tasks. Rather, the conclusion is that the variability of a person on one task is somewhat related to his variability on certain other tasks and if one is to speak of such variability for a person, one must specify the tasks on which statements are based. If more reliable means can be developed for observing intraindividual variability, better defined clusters of tasks may appear, but at present from among the tasks observed here the tasks measured by the Aiming and Number Facility tests provide the best indicators of variability.

The interpretation of these findings depends on other observations and analyses. A series of analyses of variance revealed that the test forms are not equivalent and also that a significant practice or learning effect was present, insofar as on all six of the tests daily mean scores for the group tended to increase from the beginning to the end of the experiment. On five of the tests there was no evidence that the time of day of testing was related to mean scores on the tests, but on the sixth test there was some suggestion that this relationship might exist.

Fiske raised the question regarding the relationship between the variability index and the value of the mean. The correlations between the variance index and mean index for each of the tests here were: Aiming, .41; Flexibility of Closure, .55; Number Facility, .30; Perceptual Speed, -.26; Speed of

Closure, .57; Visualization, .03. Five of the coefficients are significant; three are positive and moderately high; one is negative. Examination of two of the bivariate distributions provided no evidence that the variance indices were restricted at the low and high ends of the distributions of mean indices and the relationships appeared rectilinear. The fact that persons with high scores tended to be more variable provides some support for the hypothesis that variability extends the opportunity for the development of adaptive behavior.

The questionnaires completed by Ss at the end of the experiment suggested that they were well motivated throughout the experiment, and 83% reported that they consistently put forth all of their effort in doing as well as they could. Ninety-two percent reported that they were able to work on these tests much more effectively on some days than they could on others. Eighty-eight percent reported that on the whole they enjoyed taking the tests. The test they enjoyed least was the Aiming test and the test they enjoyed next least was the Number Facility test, the two that provided the best variance indices.

Immediately after the last form of the last test was administered, the students responded to an open-ended question asking what they thought the real purpose of the experiment was. Thirty-three percent provided the explanation given at the beginning of the experiment—to compare technology students to other students. Eighteen percent of the students reported that their perception of the purpose of the experiment was related to the consistency of behavior. Other reported purposes related to describing technology students, learning and improvement, motivation, and eye movements. Only four students reported that they did not know what the purpose of the experiment was. At the end of the questionnaire, students were presented with a checklist of five items pertaining to the purpose of the experiment. In responding to this list, 57% of the Ss checked the item, "The experiment was concerned with the consistency of my test behavior," 33% responded that the purpose was, "To determine how well I did on these tests in relation to my fellow students also taking the tests." These figures

suggest that a reasonably large proportion of the Ss had some realization that the experiment was concerned with the consistency of behavior but there was nothing to indicate that Ss for the most part were strongly motivated to behave consistently.

A supplementary analysis suggested that the indices of intraindividual variability used here are related to the predictability of students' academic behavior. For example, groups divided on the basis of the total variance index derived from the 20 forms of all of the six tests (120 scores), into high and low variability groups, did differ in predictability. The average error of grade-point prediction for the high variability group was .28, for the low variability group .009, a difference statistically significant between the .01 and .05 levels.

CONCLUSIONS

Large and significant differences are found among individuals in variability of behavior over time. For example, one S had a mean score of 43 for the 20 forms of the Number Facility test, with a standard deviation of 1.75, and another S with the same mean score had a standard deviation of 4.07. On each test large individual differences are found in variability over time.

The reliability with which these differences can be observed varies from task to task and the two tasks providing the most consistent variance index were the Aiming and Number Facility tasks. These were the two tasks that placed the students under the most stress insofar as they were the least preferred by the Ss.

Variability over time on some tasks is related to variability on other tasks, but these relationships are no more than moderate, even taking into account the relative inadequacies of the means of observation, and the highest correlation between any two of the variance indices was only about .50.

If an easily observable variability characteristic had been identified extending over

the six tests, one would face a difficult problem related to the highly speeded nature of the six tests. One then would have to determine the extent to which such a generalized intraindividual variability was related to variability in speed performance, rather than to variability over different tasks. The relatively small relationships among variances observed here might well be due to the common element of speed characterizing all of the tasks and one might be justified here in concluding only that to some extent the speed with which persons perform tasks shows some consistent intraindividual variability, quite apart from the task involved. However, all of the tasks were speeded and, if this intraindividual variability were primarily a function of variations in speed, one would expect greater consistency among variance within tasks.

The obtained results suggest that at least two of the tasks studied, Aiming and Number Facility, can provide adequate measures of intraindividual variability. The next question asked earlier by Fiske is, "With what are these individual differences in variability associated?"

REFERENCES

- BERDIE, R. F. Intra-individual variability and predictability. *Educational and Psychological Measurement*, 1961, 21, 663-676.
- FISKE, D. W. An intensive study of variability scores. *Educational and Psychological Measurement*, 1957, 17, 453-465.
- FISKE, D. W., & MADDI, S. R. *Functions of varied experiences*. Homewood, Ill.: Dorsey Press, 1961.
- HULL, C. L. Variability in amount of different traits possessed by the individual. *Journal of Educational Psychology*, 1927, 18, 97-104.
- MORAN, L. J., KIMBLE, J. P., JR., & MEFFERD, R. B., JR. Repetitive psychometric measures: Equating. *Psychological Reports*, 1964, 14, 335-338.
- MORAN, L. J., & MEFFERD, R. B., JR. Repetitive psychometric measures. *Psychological Reports*, 1959, 5, 269-275.
- WECHSLER, D. *The range of human capacities*. (2nd ed.) Baltimore: Williams and Wilkins, 1950.

(Received January 25, 1968)

COMPARATIVE RELIABILITY OF PICTURE FORM AND VERBAL FORM INTEREST INVENTORIES¹

RICHARD BLAKE²

University of Omaha

Hypothesis tested was that use of picture items in occupational interest inventories results in higher reliability than is obtained using verbal items. Verbal forms were developed to parallel published picture interest inventories. These verbal forms and the published picture forms were administered to randomly divided subgroups of high school boys and Manpower Development and Training Act men, following a test-retest design. Pearson product-moment correlations were determined for the scale scores of each group of Ss, and significance of the difference between correlations obtained by picture form and verbal form groups was then tested. A possible tendency for picture forms to yield higher reliabilities was considered not strong or consistent enough to support general claims for picture item superiority.

Authors and publishers of occupational interest inventories which use pictorial items have claimed advantages for the picture type item. Some of these expected advantages are reflected in this statement by Geist (1959):

It appears that, in principle, it is possible to construct picture items which are much less ambiguous with respect to real-life referents than are most verbal items. This decreased ambiguity should lead to higher reliability and should make higher validity possible [p. 414].

More recently and with specific reference to his own instrument, the Geist Picture Interest Inventory (Geist Inventory), Geist (1964) wrote that: "Drawings increased the reliability of the Geist Inventory and provided less ambiguous stimuli than are offered by verbal interest tests [p. 17]." This higher reliability, however, was not demonstrated.

The purpose of this study was to test the assumption that the use of picture items in interest inventories will result in higher reliability than would be obtained using verbal items. This problem is of particular importance since the most widely used interest inventories rely exclusively on verbal items.

¹ The data for this study were collected as part of a doctoral dissertation at the University of Missouri, John L. Ferguson, major advisor.

² Requests for reprints should be sent to the author, University of Omaha, 60th and Dodge, Omaha, Nebraska 68101.

PROCEDURE

An approach to testing hypotheses concerning the superiority of picture items has been suggested by Hahn (1965). In his review of the Geist Inventory, Hahn stated that:

No proof is offered that the drawings represent "stimuli which are closer to those he experiences in real life." A simple test of this assumption would have been to offer a verbal form of the test with the items being the names of the occupations, or activities, represented by the drawings [p. 1273].

In this study verbal forms were developed to parallel two published picture interest inventories. Verbal forms and picture forms were presented to Ss following a test-retest design, and the results were analyzed with reference to the relative reliability of verbal and picture forms.

Instruments

The two interest inventories used were the Geist Picture Interest Inventory and the California Picture Interest Inventory (California Inventory). The Geist Inventory has 113 drawings of vocational and avocational activities and 19 drawings of objects associated with activities. These 132 pictures are arranged in 44 triads with separate brief instructions for each triad. Examinees select one picture in each triad according to the instructions for that triad. Responses to illustrations yield scores for 11 interest areas or scales. These scales are: 1. Persuasive, 2. Clerical, 3. Mechanical, 4. Musical, 5. Scientific, 6. Outdoor, 7. Literary, 8. Computational, 9. Artistic, 10. Social Service, 11. Dramatic. The California Inventory consists of 159 pictures showing men engaged in various work activities. The inventory is divided into two parts. Part I presents the picture stimuli in 53 triads to which the examinee responds in forced-choice fashion indicating both

TABLE 1

COMPARISON OF TEST-RETEST CORRELATIONS FOR VERBAL AND PICTURE FORMS
OF SCALES FROM TWO INTEREST INVENTORIES

| Inventory | Scale | | | | | | | | | | | <i>Mdn</i> |
|--------------------------|-------|-------|------|------|------|------|-------|-----|------|------|-----|------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | |
| Geist ^a | | | | | | | | | | | | |
| Eleventh-grade boys | | | | | | | | | | | | |
| Verbal (<i>n</i> = 61) | .62 | .49 | .72 | .80 | .80 | .68 | .52 | .74 | .76 | .44 | .58 | .68 |
| Picture (<i>n</i> = 56) | .74 | .79** | .87* | .78 | .76 | .82 | .80** | .74 | .78 | .44 | .52 | .78 |
| MDTA ^b men | | | | | | | | | | | | |
| Verbal (<i>n</i> = 38) | .45 | .36 | .71 | .72 | .68 | .77 | .53 | .63 | .65 | .42 | .56 | .63 |
| Picture (<i>n</i> = 37) | .63 | .44 | .79 | .72 | .45 | .78 | .65 | .76 | .63 | .76* | .42 | .65 |
| California ^c | | | | | | | | | | | | |
| Eleventh-grade boys | | | | | | | | | | | | |
| Verbal (<i>n</i> = 58) | .73 | .86 | .64 | .87 | .75 | .83 | .70 | .84 | .66 | | | .75 |
| Picture (<i>n</i> = 51) | .86* | .92 | .84* | .94* | .87* | .85 | .86* | .83 | .73 | | | .86 |
| MDTA men | | | | | | | | | | | | |
| Verbal (<i>n</i> = 36) | .72 | .81 | .64 | .74 | .73 | .72* | .59 | .58 | .29 | | | .72 |
| Picture (<i>n</i> = 35) | .65 | .90 | .77 | .84 | .63 | .40 | .64 | .76 | .65* | | | .65 |

^a 1. Persuasive, 2. Clerical, 3. Mechanical, 4. Musical, 5. Scientific, 6. Outdoor, 7. Literary, 8. Computational, 9. Artistic, 10. Social Service, 11. Dramatic.

^b Manpower Development and Training Act.

^c California Inventory has only nine scales: 1. Interpersonal Service, 2. Natural (Outdoor), 3. Mechanical, 4. Business, 5. Esthetic, 6. Scientific, 7. Verbal, 8. Computational, 9. Time Perspective.

* $p < .05$, one-tailed.

** $p < .01$, one-tailed.

the most and least liked picture. In Part II, 30 of the pictures are repeated, this time being presented individually, with the examinee indicating either "like" or "dislike" for each picture. Responses to the illustrations yield scores for 9 scales. These are: 1. Interpersonal Service, 2. Natural (Outdoor), 3. Mechanical, 4. Business, 5. Esthetic, 6. Scientific, 7. Verbal, 8. Computational, 9. Time Perspective.

Brief verbal descriptions of the respective picture items are presented in both the Geist and the California Inventory manuals. These verbal descriptions served as a base for developing verbal items. When the verbal descriptions had been changed so as to make them appear suitable as inventory items, they were submitted to each of three judges. These judges were asked to accept or reject the verbal items on the basis that they were consistent with the author's apparent intent as evidenced by the picture content and the author's own verbal description of the item. All items were reworked until they were made acceptable to all judges. As a further check on the verbal items, matching tests were prepared with the picture items to be matched with their verbal counterparts. Each of these matching tests was given to 10 adults who had not had previous experience with the Geist or California Inventories. The average percentage of correct matching on the Geist Inventory was 99% and on the California Inventory, 98%.

Subjects and Collection of Data

The Ss for the study were eleventh-grade boys from a central Missouri high school and male students from a Manpower Development and Training

Act (MDTA) Basic Education Program in southern Missouri.

The eleventh-grade boys were randomly divided and assigned to one of two groups. The first group was tested and retested after a 3-wk. interval with the picture forms. This group was designated the Picture Form group. Each S in this group took both the Geist Inventory picture form and the California Inventory picture form on both testing and retesting. The second group was tested and retested at the same time as the Picture Form group. This second group was given the verbal forms and was designated the Verbal Form group. Each S in the Verbal Form group took both the Geist Inventory verbal form and the California Inventory verbal form on both the testing and retesting. Difference in the size of *N* reported on Table 1 results from unusable answer sheets and absences on retesting. The procedure described for eleventh-grade boys was replicated with the MDTA men with the exception that the time interval between testing and retesting was 4 rather than 3 wk. and that verbal form items were administered orally to the MDTA men because of their known reading deficiencies.

Analysis of the Data

The Pearson product-moment correlations were determined for each group of Ss. Correlations were calculated using the raw scores for each of the different scales included in the inventories. The significance of the difference between the correlations obtained by the Picture Form and Verbal Form groups was then tested using Fisher's *Z'* trans-

formation (Johnson & Jackson, 1959). This procedure was followed for both the Geist and the California Inventories and for both the eleventh-grade boys and the MDTA men.

RESULTS

Table 1 presents the findings for both the Geist and California Inventories and for both the eleventh-grade boys and MDTA men.

On the Geist Inventory, considering both eleventh-grade boys and MDTA men, the Picture Form groups had significantly higher test-retest correlations than the Verbal Form groups on 4 of the 22 comparisons made. On 7 of the comparisons made, the Verbal Form groups had the higher correlations, but on none of these did the difference reach significance.

On the California Inventory, considering both eleventh-grade boys and MDTA men, the Picture Form groups had significantly higher correlations on 6 of the 18 comparisons made. On 4 of the 18 comparisons made, the Verbal Form groups had the higher correlations, and 1 of these, Scale 6 for MDTA men, was significantly higher.

An examination of Table 1 reveals an additional finding which may or may not be directly relevant to this study but is nevertheless worth noting. On both the Geist and California Inventories there was a definite tendency for the eleventh-grade boys to obtain higher correlations than those obtained by the MDTA men. On 35 of the 40 com-

parisons that can be made between the eleventh-grade boys and MDTA men, the boys obtained the higher correlations.

CONCLUSION

Considering the results for both MDTA men and eleventh-grade boys on both the Geist and California Inventories, in all but one instance where a significant difference in reliability was found this difference favored the picture form. However, for the majority of scales the reliability of the picture forms was not significantly higher than for the verbal forms. Although there was an apparent tendency for the picture form scales to yield higher reliabilities, this tendency was neither strong nor consistent, and general claims for the superior reliability of picture items in interest inventories do not seem justified without further evidence.

REFERENCES

- GEIST, H. The Geist Picture Interest Inventory: General Form: Male. *Psychological Reports*, 1959, 5, 413-438.
- GEIST, H. *The Geist Picture Interest Inventory: Male and Female, Manual*. Beverly Hills, Calif.: Western Psychological Services, 1964.
- HAHN, M. E. Review of the 1959 edition of the Geist Picture Interest Inventory. In O. K. Buros (Ed.), *The sixth mental measurements yearbook*. Highland Park, N. J.: Gryphon Press, 1965.
- JOHNSON, P. O., & JACKSON, R. W. B. *Modern statistical methods: Descriptive and inductive*. Chicago: Rand McNally, 1959.

(Received January 26, 1968)

COMPLEX VIGILANCE: RELEVANT AND IRRELEVANT SIGNALS¹

IRWIN L. GOLDSTEIN²

University of Maryland

WILLIAM A. JOHNSTON AND WILLIAM C. HOWELL³

Ohio State University

Human monitors were required to detect additions and deletions of experimentally defined relevant signals which were presented via a computer on an 8×8 matrix display. The ratio of relevant to irrelevant stimuli on the display (20:10, 15:15, 10:20) and ratio of relevant to irrelevant signals or changes (40:20, 20:40 per 100-min. period) were investigated. Vigilance decrements were found in the detection of omit signals with the greatest decrements occurring in the experimental condition where the proportion of relevant to irrelevant signal changes was smallest.

Recent advances in sensor and computer technology require the human observer to monitor highly complex and periodically changing displays. These advances have changed the observing task that originally spurred Mackworth (1950) to use the Clock Test as a laboratory task to investigate monitoring performance. In referring to the changes in task characteristics, Kibler (1965) has suggested that the data collected in classical vigilance studies may not be applicable to contemporary monitoring problems because of changes in the signal characteristics and in the human's response requirements. Specifically, he noted that weak, brief duration signals are rarely encountered and that the human is required to monitor multiple information sources.

While the typical performance of observers in a simple display situation has been established as a monotonic decline in detected signals with increased duration of watch, there have been few definitive statements about the conditions which lead to performance decre-

ments in complex display situations like those described by Kibler. Indeed, studies investigating complex monitoring behavior did not find consistent decrements in average human performance (Adams, 1963), and those latency decrements that did occur were considered trivial (Montague, Webber, & Adams, 1965). However, recent studies (Howell, Johnston, & Goldstein, 1966; Johnston, Howell, & Goldstein, 1966) with displays more complex than those used by Adams have identified two experimental conditions which produce latency decrements: low signal frequency (the number of signals presented in a session) and high stimulus density (the average number of stimuli on the display at any given time). However, these findings were obtained under conditions in which all stimuli and signals were *relevant* and thus do not provide information on a monitoring situation where both relevant and irrelevant information are displayed.

The present study was designed to investigate just such a complex situation, that is, one in which overt responses are required to arbitrarily defined relevant signals but not to irrelevant signals. A complex format of this sort permits the exploration of some important parameters of signal frequency and stimulus density. If decrements in attention to a complex display occur only when signal frequency is low, it becomes meaningful to ask whether the important aspects of frequency are the number of overt responses (relevant

¹ This study was carried out in the Human Performance Center at the Ohio State University and was supported by the Air Force Systems Command, Research and Technology Division, Rome Air Development Center, Griffith Air Force Base, New York 13442, under Contract No. AF 30(602)-3622 with the Ohio State University Research Foundation.

² Requests for reprints should be sent to Irwin L. Goldstein, Department of Psychology, University of Maryland, College Park, Maryland 20740.

³ Now at Rice University.

signals) or the amount of stimulus change (relevant and irrelevant signals) occurring per unit time. Similarly, it becomes important to learn whether the critical aspect of stimulus density is the total number of stimuli or just the number of relevant stimuli.

METHOD

Subjects

The Ss were 48 male undergraduate students divided equally into six experimental conditions. Practiced Ss were used to minimize confounding of learning and vigilance efforts.

Apparatus and Stimulus Material

All programming of input and scoring of responses were performed using a combined IBM 1401/7094 digital computer system. Four cathode-ray-tube display consoles were linked directly to this system in such a way that four Ss could serve simultaneously under independent experimental conditions. The display consisted of a 5-in. sq. 8×8 matrix. Each of 64 cells contained either a relevant letter trigram, an irrelevant one, or nothing. Each trigram contained either an A or an E as the middle letter with the remaining letters chosen randomly from the rest of the alphabet. For one half the Ss the A trigrams were relevant, and for the other half the E trigrams were so designated. A signal was defined as either the appearance of a new trigram in a previously empty cell (*add* signal) or the disappearance of an old one (*omit* signal). If the change involved a relevant stimulus, it was a relevant signal; otherwise, it was irrelevant.

Experimental Design and Signal Programming

At any given time there were approximately 30 stimuli on the display, and in the course of a 100-min. watch period 60 changes (signals) occurred involving this information. Using this scheme, two variables in a factorial design were investigated: ratio of relevant to irrelevant stimuli on the display (20:10, 15:15, 10:20) and ratio of relevant to irrelevant signals or changes (40:20 per 100-min. period, and 20:40). In order to maintain the average density values, adds and omits occurred with equal frequency over a session. The other characteristics of the signals were determined on a random basis including: the selection of the first and third letter of each trigram from a pool of all possible combinations; the specific stimuli added and omitted; the order of occurrence of adds and omits; and the selection of cells for placement of signals with the obvious restriction that the cell had to be empty for a signal to be added and had to be occupied for a signal to be omitted. Only one stimulus could occupy a cell at any given time and no more than one could change at once. The intersignal intervals varied normally around a mean of 100 sec. and standard deviation of 10 sec. with the restriction that all values were multiples of 10 sec.

Procedure

Each S completed a 100-min. practice and a 100-min. experimental session. The 48 Ss were divided into the six experimental conditions. Four Ss performed simultaneously but were under different experimental conditions. The Ss were instructed to remain seated, quietly but awake, for the entire 100 min. of each session. With no exceptions Ss followed their instructions faithfully.

The Ss were informed that stimuli might either appear or disappear and that they were to respond only to relevant signals by pushing immediately a detect button beneath their left hand. The Ss were also required to illuminate the cell in which the signal occurred with a light pencil held in the right hand. This latter procedure prevented S from using the detect button indiscriminately and also permitted the authors to match up detections and signals. One important characteristic of this task was the relative persistence of signal states. When a signal occurred in any cell, the new state of that cell (i.e., signal present or absent) remained in effect indefinitely, subject only to the rules of random selection which governed the choice of any cell for change. If S detected a response but could not locate where it occurred, he could push the DNO (Don't Know) button. These kinds of partial responses were very rare. The computer provided immediate confirmation to S after each step in the response sequence so that S could determine if his responses were properly recorded.

RESULTS AND DISCUSSION

An earlier study (Howell et al., 1966) found that the monitoring behaviors for adds and omits are quite distinct, seemingly involving different search and memory processes. Furthermore, detection of omits depends on short-term memory to a greater extent than does detection of adds. Therefore, the data for the two signal types were analyzed separately. For a given density condition, the signals of each type were organized into two blocks of 10 signals each for the 40:20 frequency condition, and into two blocks of 5 signals each for the 20:40 frequency condition. An ordinal time scale in the form of signal blocks was used in order to avoid possible artifactual latency measurements produced by an uneven distribution of the relatively small number of signals across time blocks. The relative persistence of signals permitted the examination of the effects of the independent variable on detection latency without setting an arbitrary upper boundary which would limit all scores to a certain range. The median latency scores were chosen as the most sensi-

tive measure of individual monitoring performance because these scores were not affected by extensively long latencies which occasionally occurred under all conditions.

An analysis of variance of the data for adds revealed that none of the experimental sources of variance was significant ($p > .05$). This finding, contrasted with the significant effects found with the omit signals (discussed below), points to the complicated processes underlying monitoring behavior with complex displays. The present data support earlier statements (Johnston et al., 1966) that there are very different search and storage operations for add and omit signals. In this particular case the display was filled with an average of 30 different trigrams which makes storage operations very difficult. Therefore, S was dependent upon a continual search pattern which appears to be very sensitive to add signals. The S s actually detected 91% of the add signals compared with only 81% of the omit signals. This has led to speculation that add signals are somewhat more attention demanding than omit signals,⁴ and that this superiority, whatever its source, apparently obscured the effects of the independent variables on add signals. More experiments with emphasis on search and storage operations are planned on this apparent difference in sensitivity to adds and omits.

An analysis of variance on the latency scores for omit signals indicated that: the signal ratio or frequency effect was significant ($p < .01$); latencies increased significantly over blocks ($p < .05$); and the Blocks \times Frequency interaction just missed significance ($F = 4.07$ was required for $p < .05$ and $F = 3.93$ was obtained). The density variable and all of its interactions were not significant ($p > .05$).

As can be seen from an examination of Table 1, these data reveal that there was an overall decrement in performance for omit signals. Also, it is apparent that the latency scores for the 20:40 relevant/irrelevant frequency condition are higher than for the 40:20 condition. Table 1 also indicates that the main

TABLE 1

MEDIAN DETECTION LATENCY SCORES FOR TWO SIGNAL BLOCKS OBTAINED FOR THE ADD AND OMIT SIGNALS UNDER 20:40 AND 40:20 RELEVANT/IRRELEVANT FREQUENCY

| Signal | Block | |
|--------|-------|-------|
| | 1 | 2 |
| Omit | | |
| 20:40 | 9.40 | 14.11 |
| 40:20 | 6.97 | 7.67 |
| Add | | |
| 20:40 | 5.31 | 5.24 |
| 40:20 | 4.06 | 4.84 |

decrement occurred in the 20:40 condition. This was supported by a Newman-Keuls test on the Frequency \times Block interaction which showed that Block 2 was significantly different from Block 1 for the 20:40 condition ($p < .05$) but not for the 40:20 condition ($p > .05$). The analysis also revealed that the scores for the 20:40 and 40:20 conditions were significantly different at Block 2 ($p < .05$) but not at Block 1 ($p > .05$).

The present data for omit signals support Jerison and Pickett's (1964) reinforcement theory which assumes that signals are reinforcers, that the monitoring task is composed of observing responses, and that performance improves with increases in the percentage of observing responses reinforced. Also, these data appear to indicate that lack of a performance decrement with high frequency is not accounted for by the amount of stimulus change occurring per unit time as the novelty theory might predict, but rather the reinforcement of observing responses to relevant stimulation.

The lack of any effects involving stimulus density is puzzling. It is possible that the critical aspect of density is the total number of stimuli on the display (which was the same for all conditions) rather than the number of relevant stimuli. It is also possible that there were two offsetting processes that obscured any density effect. As you change from a 10:20 density to a 20:10 density, you increase the number of stimuli to be stored which should inhibit performance; but you also decrease the amount of irrelevant stimula-

⁴ Though the physical characteristics of the equipment do not account for this concept, subjective reports from S s and E s support this view.

tion on the display which should improve scanning behavior and thus enhance performance.

In summary, the present study examined the performance of human Ss in monitoring complex displays. Decrements to omit signals were observed with the main source of the decrements being the 20:40 frequency condition. It was under these conditions that monitors received the smallest number of reinforcing signals and the largest amount of extraneous stimulation. The effects of density and add versus omit signals require further work before any systematic statements can be offered about their effects upon human performance.

REFERENCES

- ADAMS, J. A. Experimental studies of human vigilance. *United States Air Force ESD Technical Documentary Report*, 1963, No. 63-320.
- HOWELL, W. C., JOHNSTON, W. A., & GOLDSTEIN, I. L. Complex monitoring and its relation to the classical problem of vigilance. *Journal of Organizational Behavior and Human Performance*, 1966, 1, 129-150.
- JERISON, H. J., & PICKETT, R. M. Vigilance: The importance of the elicited observing rate. *Science*, 1964, 143, 970-971.
- JOHNSTON, W. A., HOWELL, W. C., & GOLDSTEIN, I. L. Human vigilance as a function of signal frequency and stimulus density. *Journal of Experimental Psychology*, 1966, 72, 736-743.
- KIBLER, A. W. The relevance of vigilance research to aerospace monitoring tasks. *Human Factors*, 1965, 7, 91-99.
- MACKWORTH, N. H. Researches on the measurement of human performance. *Medical Research Council Special Report Series* (England), 1950, No. 268.
- MONTAGUE, W. E., WEBBER, C. E., & ADAMS, J. A. The effects of signal and response complexity on eighteen hours of visual monitoring. *Human Factors*, 1965, 7, 163-172.

(Received January 26, 1968)

PREDICTION OF JOB SUCCESS FOR HOSPITAL AIDES AND ORDERLIES FROM MMPI SCORES AND PERSONAL HISTORY DATA¹

JAMES N. McCLELLAND² AND FEN RHODES

California State College at Long Beach

The MMPI and selected personal history data were evaluated as predictors of job success for hospital aides and orderlies. Thirteen individual performance measures and a weighted composite criterion were employed for each job. All but 1 of the 10 basic MMPI scales and 10 of 13 biographical predictors showed small correlations with at least 1 of the individual criterion measures. Corrected multiple *R*s of .48 were obtained between each weighted composite criterion and combined test and biographical predictors. The biographical data proved relatively more useful in this prediction. Factor analyses of individual criterion measures yielded similar factor structures for both jobs.

The problem of selection of nurses' aides and orderlies in hospitals has become increasingly important in recent decades. Not only do these auxiliary personnel comprise a significant proportion of the hospital staff, but also high turnover rates within their ranks and inefficient performance on the job make the work they perform more costly than necessary. Proper selection of individuals to fill these positions is, therefore, of considerable importance. It was the purpose of this study to investigate the usefulness of the MMPI and biographical data from application forms in selecting these personnel. Criteria for judging adequacy of selection methods were various performance ratings and attendance and tenure data.

Both MMPI and autobiographical data have been used successfully for the selection of personnel in hospitals and in industry. Research most clearly related to the present study includes use of the MMPI by Rowe (1957) to predict success of psychiatric aides, Hovey

(1956) to predict success in the clinical training of student nurses, Butterfield and Warren (1962, 1963) to predict performance and job tenure of psychiatric aides, and Bessent and Gloye (1967) to predict job performance and washout rate of mental hospital technicians.

Autobiographical data, collected by means of job application forms, have been used widely in industry to predict both job performance and job tenure. Recent representative studies include use of such data by Scollay (1957) to predict success of assistant district sales managers, by Kirchner and Dunnette (1957) to reduce turnover in a variety of office jobs, by Fleishman and Berniger (1960) to predict tenure of office workers at Yale University, by Walther (1961) to predict performance and turnover among secretaries, code clerks, and mail and record clerks, and by Buel (1965) to predict creativity of research personnel.

METHOD

Subjects

The Memorial Hospital of Long Beach, California, is a private 550-bed hospital. Since 1961 the hospital has administered the MMPI to all its nurses' aides and orderlies, either before or immediately after employment. The total pool of individuals thus made available for the study comprised 111 nurses' aides and 100 orderlies. Of these, 37 aides and 23 orderlies were still employed at the time of the study.

All aides were females and all orderlies were males. In addition to the difference in sex, the following distinctions between the two occupations were

¹ The authors gratefully acknowledge the assistance of Betty J. Greer, Director of Personnel, Memorial Hospital of Long Beach, in planning the study and gathering data. This work was accomplished under United States Public Health Service Grant No. 1-R03 MH 11910-10A1 (MSM). Computing assistance was obtained from the Health Sciences Computing Facility, University of California, Los Angeles, sponsored by National Institute of Health Grant FR-3.

² Requests for reprints should be sent to James N. McClelland, Department of Psychology, California State College, 6101 East Seventh Street, Long Beach, California 90804.

TABLE 1

MEANS OF BIOGRAPHICAL PREDICTOR VARIABLES

| Predictor variable | Aides ^a | Orderlies ^b |
|--|--------------------|------------------------|
| Age | 37.4 | 23.0 |
| Marital status (proportion) | | |
| Single | .15 | .78 |
| Married | .49 | .20 |
| Divorced | .26 | .02 |
| Widowed | .10 | .00 |
| No. dependents | 1.17 | .24 |
| Yr. education | 11.3 | 12.3 |
| Proportion with health impairment | .22 | .18 |
| Average tenure on past jobs ^c | 29.3 | 14.0 |
| Related experience ^c | 23.1 | 15.9 |
| Salary difference: Present versus previous job (dollars per mo.) | 9.39 | -14.39 |
| Proportion with restriction on hr. available for duty | .49 | .39 |
| Length of time local resident ^c | 38.8 | 47.8 |

^a *N* = 72.^b *N* = 54.^c In months.

deemed significant enough to keep the groups separate throughout most of the study:

1. Nurses' aides are expected to give close support and attention to a small number of patients on a single ward or floor, whereas orderlies have a much broader assignment that takes them throughout the hospital.

2. Only orderlies perform certain functions, such as catheterizations for male patients.

3. Aides make decisions that are more crucial, and, hence, they were expected to exhibit judgment and initiative at a higher level than are orderlies.

4. Because their tasks are more varied, orderlies have more responsibility for scheduling their work.

Job-performance data and biographical data were not complete for all *Ss* in the pool. Consequently, the number of cases employed varied for different parts of the study. However, the number of *Ss* in either classification, aides or orderlies, available for any one part of the study was never less than 54. Actual numbers of individuals employed for each part of the study are indicated in the body of the report.

Predictor Variables

Scores for a total of 18 MMPI scales were tested for validity as predictors. Scales used were the basic 10 (*Hs*, *D*, *Hy*, *Pd*, *Mf*, *Pa*, *Pt*, *Sc*, *Ma*, *Si*), plus *L*, *F*, *K*, and five scales corrected by the *K* factor (*Hs* + *K*, *Pd* + *K*, *Pt* + *K*, *Sc* + *K*, *Ma* + *K*). The following biographical information was obtained from application blanks (Table 1):

1. Age at time of employment. Ages ranged from 18 to 57 for aides and from 18 to 51 for orderlies.

2. Marital status. Recorded as 1 or 0 for single versus not single, married versus not married, divorced or separated versus not, and widowed versus not widowed.

3. Number of dependents.

4. Education in years.

5. Health impairment. Recorded as absence (0) or presence (1) of disability other than eyeglasses.

6. Average tenure (months) on past jobs.

7. Related experience. Recorded in years for persons reporting previous jobs in related fields.

8. Salary difference. Computed as present minus last previous job in dollars per month.

9. Restriction on hours available for duty. Recorded as available on all shifts (0) or only at specified times (1).

10. Number of months resident of local area.

Approximately half of those in the aide group had taken the MMPI before employment and half shortly afterwards. Since the potential effect on test scores caused by this difference in administration is well known (e.g., Green, 1951), it was deemed necessary to test for differences in responses made by *Ss* who had taken the MMPI as applicants and those who had taken it as employees. A *t* test for significance of differences between mean scores showed no difference ($p < .05$) on any of the 18 scales. Consequently, no distinction was made between aides who took the MMPI before and those who took it after employment.

Criterion Measures

Criterion data for which prediction was sought comprised supervisory ratings on a series of 5-point performance scales plus data on attendance and job tenure. The ratings, completed at normal review periods, were described on the rating form as: (a) quality of work (accuracy), (b) volume of work, (c) ability to follow directions (instructions), (d) acceptance of responsibility (pressure, initiative, judgment), (e) effective use of time, equipment and supplies, resourcefulness, ability to organize, (f) personal grooming (hygiene, appearance, appropriateness of dress), (g) attendance record (tardiness, absenteeism), (h) observance of confidential nature of patient and hospital business, and (i) relationship with others (co-workers, patients, public, visitors), courtesy, cooperativeness.

Other criterion data used were number of months employed by the hospital at the time of the study or at termination, average number of hours absent from work per month, average number of absences per month, and presence or absence of recorded instances of aberrant behavior detrimental to job performance.

In addition to the individual criterion measures described above, a weighted composite criterion was developed to provide a single measure of overall job performance. To arrive at appropriate weights for the different aspects of job performance for which quantitative data were available, 20 supervisors were asked to distribute a total of 100 points among 10 selected job-performance characteristics based upon the perceived importance of each characteristic with respect to overall job performance. Aide and orderly positions were evaluated

separately by the 20 supervisors. There was a minimum interval of 1 wk. between rating of the first and second jobs, and supervisors were not told at the beginning that they would even be rating a second job. Carry-over effects still remaining were counterbalanced by having half of the supervisors assign weights first for aides and the other half first for orderlies. Average weights assigned for each of the 10 performance characteristics included in the composite criterion are shown in Table 2. The weights for hours absent and frequency of absences were changed to negative values. Weights were multiplied by individual ratings for each employee expressed in standard score form and the products summed to provide a composite job-performance score.

Data Analysis

Zero-order correlations between all of the MMPI scales and each of the 13 job-performance criteria were obtained for both aides and orderlies. Correlations between biographical data items and individual criterion measures were similarly computed for both job groups. In addition all test and personal history predictors were correlated with the weighted composite criterion measures developed for aides and orderlies on the basis of supervisory judgments.

Multiple linear regression analyses were performed for each of the job groups using in turn the MMPI scales and the biographical data items as predictors of each of the 13 individual job-performance measures. Composite weighted criterion performance was predicted for each job group by a linear regression analysis employing the entire set of test and non-test predictor variables.

Finally, separate principal-components factor analyses were carried out for aides and orderlies using 13 of the individual criterion measures for the purpose of identifying the number and types of independent factors operating in observed performance on these two jobs. A Varimax rotation was performed on the obtained matrices.

RESULTS

Zero-order correlations between MMPI scores and the 13 job-performance measures were generally small, and most were statistically insignificant ($p > .05$). Fourteen percent were significant for aides, with a median significant r of .25. For orderlies only 9%

³ Tables for both job groups giving zero-order and multiple correlations of all predictors with individual criteria have been deposited with the National Auxiliary Publications Service. Order Document No. 0123 from National Auxiliary Publications Service of the American Society for Information Science, c/o CM Information Sciences, Inc., 22 West 34th Street, New York, New York 10001. Remit in advance \$3.00 for photocopies or \$1.00 for microfiche and make checks payable to: Research and Microfilm Publications, Inc.

TABLE 2
RATED IMPORTANCE OF DIFFERENT ASPECTS OF JOB PERFORMANCE FOR AIDES AND ORDERLIES

| Performance characteristic | Assigned weight ^a | |
|------------------------------|------------------------------|-----------|
| | Aides | Orderlies |
| Hr. absent | -4 | -3 |
| Frequency of absences | -4 | -5 |
| Quality of work | 22 | 21 |
| Volume of work | 14 | 13 |
| Ability to follow directions | 17 | 16 |
| Acceptance of responsibility | 10 | 9 |
| Resourcefulness | 11 | 11 |
| Personal grooming | 6 | 7 |
| Relationship with others | 10 | 12 |
| Tenure | 2 | 3 |

^a Represents average of ratings by 20 supervisors familiar with both jobs.

were significant, with a median of .26. All criterion measures but three, however, could be significantly predicted at least minimally from one or another MMPI score. The three exceptions were tenure, attendance as rated by supervisors, and relationship with others (also evaluated by supervisors), none of which showed significant correlations with MMPI predictors for either job category. In addition, for orderlies, rated discretion of employees regarding patient and hospital business could not be predicted from the MMPI. The objective measures of absenteeism (i.e., hours absent and frequency of absences) were among the most predictable criterion variables for both groups and were the most predictable for aides. Of the 18 MMPI scales, only *Si* failed to correlate significantly with any criterion measure for either job group.

Relationships between biographical data predictors and individual performance criteria were somewhat larger compared with those for the MMPI (median significant r of .28 for aides, .34 for orderlies), but the total number of significant ($p < .05$) correlations in each job category was considerably reduced (6% for aides, 4% for orderlies). Two criterion measures, namely, ability to follow directions and relationship with others, could not be significantly predicted in either job group from biographical data. Hours absent was correlated significantly for aides with being divorced (.28) and with health impair-

TABLE 3
CORRELATIONS BETWEEN ALL PREDICTORS AND
COMPOSITE CRITERION

| MMPI predictor | Aides | Orderlies | Biographical predictor | Aides | Orderlies |
|----------------------|-------|-----------|--|-------|-----------|
| <i>L</i> | -.23* | .08 | Age | .02 | .24 |
| <i>F</i> | -.11 | .09 | Single | .15 | .08 |
| <i>K</i> | -.23* | .05 | Married | .22 | -.01 |
| <i>Hs</i> | -.02 | .14 | Divorced | -.28* | -.23 |
| <i>D</i> | -.12 | .20 | Widowed | -.15 | .00 |
| <i>Hγ</i> | -.30* | -.01 | No. dependents | -.18 | .12 |
| <i>Pd</i> | -.18 | -.10 | Education | .13 | .14 |
| <i>Mf</i> | .04 | .10 | Health impairment | -.02 | .17 |
| <i>Pa</i> | -.08 | -.02 | Average tenure on past jobs | .16 | .23 |
| <i>Pt</i> | .11 | .06 | Related experience | -.08 | .19 |
| <i>Sc</i> | .01 | -.02 | Salary difference: Present versus previous job | -.15 | -.03 |
| <i>Ma</i> | .01 | -.22 | Restriction of hr. available for duty | .12 | .01 |
| <i>Si</i> | .07 | .16 | Length of time local resident | -.01 | -.28* |
| <i>Hs</i> + <i>K</i> | -.21 | .18 | | | |
| <i>Pd</i> + <i>K</i> | -.28* | -.05 | | | |
| <i>Pt</i> + <i>K</i> | -.12 | .14 | | | |
| <i>Sc</i> + <i>K</i> | -.20 | .04 | | | |
| <i>Ma</i> + <i>K</i> | -.03 | -.20 | | | |

Note.—Aides, *N* = 72; Orderlies, *N* = 54.
* *p* < .05.

ment (.37). Frequency of absences was correlated significantly with being single (–.24) or divorced (.23), with number of dependents (.29), and with health impairment (.35). For orderlies, however, neither absenteeism measure correlated significantly with any biographical predictor. Three biographical predictors, namely, widowed, education, and restriction on hours available for duty, did not correlate significantly for aides or orderlies with any job-performance measure.

Correlations of both sets of predictor vari-

TABLE 4
MULTIPLE CORRELATION BETWEEN ALL PREDICTORS
COMBINED AND COMPOSITE CRITERION

| Group | Obtained <i>R</i> | Corrected <i>R</i> | Predictors included* |
|------------------------|-------------------|--------------------|---|
| Aides ^b | .52* | .48 | <i>Hγ</i> (–2.6), divorced (–58.5), widowed (–81.3), past tenure (1.0), <i>K</i> (–4.1) |
| Orderlies ^c | .54* | .48 | Local residence (–0.1), education (35.6), age (12.7), divorced (–390) |

Note.—Relative size of raw-score weight from regression analysis appears in parentheses after each scale.
* In order of relative contribution to total prediction.
^b *N* = 72.
^c *N* = 54.
* *p* < .05.

ables, that is, MMPI and biographical data, against the composite weighted criterion developed for aides and that developed for orderlies are listed in Table 3. From the total number of 31 predictors, only 5 for the aides and 1 for the orderlies correlated significantly (*p* < .05) with composite measured job performance.

Table 4 shows the multiple correlations obtained when MMPI scores and biographical data were combined as predictors of the composite performance measures for aides and orderlies. In calculating these *R*s (and those in NAPS Tables C and D) a stepwise multiple repression procedure was employed, which began with the best single predictor variable and successively selected additional variables so that the maximum increase in predictive power would be realized as each new variable was brought in. The process was terminated in every case when the *F* ratio computed for a predictor to be added next failed to reach significance at the .25 level, indicating that inclusion of further variables would not meaningfully improve the multiple *R* obtained thus far. All predictor variables passing this criterion are listed in the tables. The corrected or “shrunk” *R* shown for each obtained *R* is an unbiased estimate of

TABLE 5
FACTOR MATRIX FOR 12 JOB-PERFORMANCE
MEASURES: AIDES

| Performance measure | Factor | | | |
|---|--------|------|------|------|
| | I | II | III | IV |
| Hr. absent | -.07 | .88 | -.15 | -.05 |
| Frequency of absences | -.17 | .88 | -.11 | -.12 |
| Aberrant behavior | -.09 | -.09 | -.43 | -.13 |
| Quality of work | .49 | -.18 | .25 | .64 |
| Volume of work | .59 | -.21 | .18 | .53 |
| Ability to follow directions | .69 | -.15 | .07 | .16 |
| Acceptance of responsibility | .43 | -.16 | .31 | .64 |
| Resourcefulness | .52 | -.10 | .16 | .52 |
| Personal grooming | .15 | -.13 | .45 | .49 |
| Attendance (rated) | .44 | -.55 | .04 | .28 |
| Discretion re patient and hospital business | .42 | -.35 | .10 | .22 |
| Relationship with others | .65 | -.12 | .26 | .33 |

Note.—Factors shown are orthogonal and were obtained using a principal-axes solution with Varimax rotation. *N* = 90.

the population correlation. As such, it approximates the size R that would be obtained if the regression weights derived for the present employee sample were applied to a second, cross-validation sample.

Tables 5 and 6 show the results of the factor analysis conducted for each job of the individual criterion measures used in the study. Tenure was not included in these analyses, although it was used as a criterion variable in other parts of the study, because this variable was assigned for both aides and orderlies a very low weight by supervisors relative to the judged importance of other aspects of job performance (Table 2). It was considered undesirable and possibly misleading to include in this instance a variable rather clearly established as having little bearing on how well an incumbent performs a job. Factors were extracted by the method of principal components, using estimated communalities (R^2 s) in the diagonal of each matrix. Orthogonal rotations were performed by the Varimax technique. Factors retained for rotation accounted for at least 1% of the total variance.

DISCUSSION

In general, findings of this study are similar to many in which a sampling of biographical data, a personality measure, or both, have been used to predict job-performance criteria. For example, the corrected multiple R s of .48 reported here (Table 4) are very close to the average multiple R s of approximately .50 obtained by Rowe (1957), who used biographical and MMPI data to predict a composite job-performance criterion for psychiatric aides. Using several individual biographical predictors combined on a logical basis, Scollay (1957) obtained biserial correlation coefficients of .32 and .23 in predicting managerial success expressed on a dichotomous scale.

Various studies tend to reveal some consistency in the relative effectiveness of the MMPI scales in predicting job performance. In this study the scales most often found useful as predictors both of individual and composite criteria were Hs , Pd , Mf , Ma , and F . On the other hand Butterfield and Warren (1963) found the K , Pd , and Ma scales most

TABLE 6
FACTOR MATRIX FOR 12 JOB-PERFORMANCE
MEASURES: ORDERLIES

| Performance measure | Factor | | | | |
|---|--------|------|------|------|------|
| | I | II | III | IV | V |
| Hr. absent | -.08 | .78 | .02 | -.02 | -.15 |
| Frequency of absences | -.06 | .78 | .02 | -.13 | -.18 |
| Aberrant behavior | -.01 | .18 | -.11 | -.12 | -.41 |
| Quality of work | .74 | -.19 | .13 | .15 | .17 |
| Volume of work | .84 | -.06 | .25 | .06 | -.06 |
| Ability to follow directions | .40 | -.12 | .60 | .16 | .10 |
| Acceptance of responsibility | .76 | -.22 | .28 | .14 | .15 |
| Resourcefulness | .72 | .00 | .31 | .27 | -.20 |
| Personal grooming | .40 | -.10 | .54 | -.08 | .26 |
| Attendance (rated) | .35 | -.56 | .31 | .24 | -.06 |
| Discretion re patient and hospital business | .16 | -.13 | .05 | .46 | .13 |
| Relationship with others | .43 | -.32 | .43 | .25 | .07 |

Note.—Factors shown are orthogonal and were obtained using a principal-axes solution with Varimax rotation. $N = 86$.

useful, and Bessent and Gloye (1967) found the Pd , Mf , Pa , and Ma scales best. No doubt differences in outcomes of success employed in the job requirements are responsible for differences of this kind.

Particularly interesting findings of the present study were:

1. The K scale alone was a useful predictor for three criterion variables for aides (viz. quality of work, ability to follow directions, and resourcefulness), but was of no significance in predicting the performance of orderlies.

2. The F scale was significantly correlated with absenteeism for both aides and orderlies. However, absenteeism was positively correlated with F for aides and negatively for orderlies. No explanation is offered for this finding.

3. Every one of the 10 basic MMPI scales except the Si scale was significantly correlated with at least one of the individual criterion variables. Application of the K suppressor

variable to five of the scales did not materially increase their predictive efficiency except for the *Pd* scale. For this scale the number of criterion variables predicted at a significant level was increased from two to seven for orderlies when *K* was applied.

4. Divorced status yielded the highest correlation (.50) with an individual criterion of any of the predictor variables, either biographical or MMPI. This variable also correlated with five other criterion measures, making it the most useful biographical predictor and among the most useful of all predictors.

5. Education was not significantly correlated with any individual job-performance measure. This occurred despite the fact that the amount of education in the sample studied ranged from 8 to 14 yr. However, it was noted that, in predicting the composite performance criterion for orderlies, education did appear as one of the most useful variables.

6. Although there was a very great range of ages represented in the sample, this variable did not appear to be a useful predictor, except that older orderlies were rated as more resourceful.

7. Biographical variables were more important than MMPI scores for predicting composite job performance. The *Hy* and *K* scales were the only MMPI variables included in the multiple regression prediction, and these applied to aides only.

The factor analyses (Tables 5 and 6) revealed similar structures for both types and jobs. Four orthogonal factors were extracted for the aides and five for orderlies. These factors were named as follows: *Aides*—I. General rater bias (halo effect) reflecting adequacy of overall job performance, II. Absenteeism, III. Personal adjustment, IV. Initiative. *Orderlies*—I. General rater bias (halo effect) reflecting adequacy of overall job performance, II. Absenteeism, III. Personal relationship with supervisor, IV. Discretion in dealing with patients, V. Personal adjustment.

The appearance of a "resourcefulness and initiative" factor specific to aides, and a "personal relationship with supervisor" factor specific to orderlies is consistent with the differences described earlier between these two

jobs. On the one hand, aides are more directly and immediately involved in emergency problems with patients and thus have the opportunity to be resourceful and to demonstrate initiative. On the other hand, each orderly works simultaneously for several nursing supervisors, a situation that renders the ability to establish satisfactory personal relationships with others especially important.

It is emphasized that the results of this study must be interpreted in light of the fact that the sample studied was a group of present employees and not unselected applicants. Had the study been based on the latter group, correlations of the various predictors with performance criteria would have been moderately greater probably without exception.

REFERENCES

- BESSENT, T., & GLOYE, E. An analysis of MMPI records for mental hospital technicians. Paper presented at the meeting of the Western Psychological Association, San Francisco, May 1967.
- BUEL, W. D. Biographical data and the identification of creative research personnel. *Journal of Applied Psychology*, 1965, **49**, 318-321.
- BUTTERFIELD, E. C., & WARREN, S. A. The use of the MMPI in the selection of hospital aides. *Journal of Applied Psychology*, 1962, **46**, 34-40.
- BUTTERFIELD, E. C., & WARREN, S. A. Prediction of attendant tenure. *Journal of Applied Psychology*, 1963, **47**, 101-103.
- FLEISHMAN, E. A., & BERNIGER, J. One way to reduce office turnover. *Personnel*, 1960, **37**, 63-69.
- GREEN, R. F. Does a selection situation induce testees to bias their answers on interest and temperament tests? *Educational and Psychological Measurement*, 1951, **11**, 503-515.
- HOVEY, H. B. Positive characteristics of MMPI elevations in a nursing group. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: University of Minnesota Press, 1956.
- KIRCHNER, W. K., & DUNNETTE, M. D. Applying the weighted application blank technique to a variety of office jobs. *Journal of Applied Psychology*, 1957, **41**, 206-208.
- ROWE, F. B. The selection of psychiatric aides: Criterion development and prediction. (Doctoral dissertation, University of Maryland) *Dissertation Abstracts*, 1957, **17**, 2674-2675.
- SCOLLAY, R. W. Personal history data as a predictor of success. *Personnel Psychology*, 1957, **10**, 23-26.
- WALTHER, R. H. Self-description as a predictor of success or failure in foreign service clerical jobs. *Journal of Applied Psychology*, 1961, **45**, 16-21.

(Received February 1, 1968)

SKIMMING LISTS OF FOOD INGREDIENTS PRINTED IN DIFFERENT SIZES¹

E. C. POULTON²

Applied Psychology Research Unit, Cambridge, England

72 housewives aged 22-64 yr. searched for particular words in lists of ingredients printed in 10-, 7.5-, 6-, and 4-pt. Univers lower case type with 10% leading. There were 4 sets each of 15 lists, and a 4×4 factorial design was used confounding order and sets of lists. Illumination levels of 40 and 2 ftc. were tried in separate experiments. There was a large ($p < .001$) drop in the rate of locating ingredients when the size of print was reduced from 6 to 4 pt. Increasing the size from 6 to 10 pt. had less effect ($p < .05$). In the dim light, 4 housewives over 50 yr. could not locate any ingredients printed in 4 pt. The conclusion was that ingredients should not be printed in lower case Univers smaller than about 6 pt., which has the same x height or apparent size as 8-pt. book type.

It is clear from common observation that housewives do not read through the complete list of ingredients on each package of food which they buy. If they look at all, they probably search the list to see only if the food contains an ingredient they particularly want it to contain, or perhaps an ingredient which they do not want it to contain. Assuming this to be the case, the clarity of print for lists of ingredients would be tested most appropriately by asking housewives to search the lists for particular ingredients. This is the method which has been used here.

Small food containers have sufficient surface available only for small amounts of print. If lists of ingredients are to be printed in full, the print may have to be minute. One aim of the present experiments was to determine the smallest size of print which housewives could be expected to read. It was hoped to be able to specify a minimum size of print,

below which there would be a marked fall in the rate at which ingredients could be read.

Considering the large amount of work which has been carried out upon the clarity of print (Burt, 1959; Luckiesh & Moss, 1942; Tinker, 1963), it might be presumed that the minimum size of print for food containers could be determined by reference to existing experimental data. However a thorough search of the literature indicated that this was not so. Most of the work has been concerned more with the optimal size of print than with the minimal acceptable size.

Luckiesh and Moss (1942, Table XVII) compared 6-pt. Textype set 46 characters to the line with 1-pt. leading with 8-, 10-, and 12-pt. set 52 characters to the line with 2-pt. leading. They used blink rate as their criterion of legibility and found practically no differences. If anything, the 8-pt. Textype was the most legible by this criterion. However Tinker (1946) has shown since that the rate of blinking is not a valid measure of difficulty in reading. In Luckiesh and Moss' experiments rate of reading hardly varied with legibility, presumably because their readers knew that they were not to be tested for comprehension.

Burt (1959, p. 12) compared sizes of Times Roman ranging only from 8 to 14 pt. Unfortunately he did not tabulate his experimental data nor indicate what statistical tests, if any, he carried out. He simply stated that 10-pt. prose was the most legible. Most

¹ This research was carried out at the request of The Metal Box Co. Ltd., which kindly supplied the printed materials. The British Food Manufacturers' Federation kindly defrayed the cost of the housewives. The author is also grateful to I. Harris and P. Harris of The Metal Box Co. for their help and encouragement. P. M. E. Altham kindly advised on the experimental design. Financial support from the British Medical Research Council is also gratefully acknowledged.

² Requests for reprints should be sent to the author, Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

students found 9 pt. equally legible, while older persons often did best with 11 or even 12 pt.

Tinker and Paterson also were concerned principally with the optimum size of print. However, in their (1932) study 6 $\frac{3}{4}$ pt. Mergenthaler's Ionic linotype on a 7-pt. body was reduced by the planographic offset-printing process. Compared with a control group, they found a 2% reduction in the students' mean rate of reading when the size of the type was reduced to 80%. Reducing the size to 50% reduced the mean rate of reading by 11%, a fall which was reliable statistically. But the catastrophic fall in rate of reading was shown only by a reduction in the size of the type to 30%. At this point the mean rate of reading had fallen by 74%; 15 of 90 students scored zero. On this evidence 3.5-pt. newspaper print should be just acceptable for students.

Unfortunately students are not comparable to housewives, since most are young. Set against this, newspaper type is designed to be read in its existing size; it is not designed to be reduced in size. Normally a revised set of letters has to be designed for every change of about 2 in point size, since smaller letters require relatively thinner lines and larger spaces within the letters ("counters"). If this holds for Mergenthaler's Ionic linotype, it is possible that a properly designed 3-pt. type might have been as acceptable for students as the 6 $\frac{3}{4}$ -pt. linotype reduced to 50%.

In order to produce data like Tinker and Paterson's, but relevant to the printing of ingredients on food containers, it was decided to use (a) housewives with a wide range of ages, rather than students, and (b) a typeface like Univers which was designed to be reduced photographically.

METHOD

Materials

Lists of ingredients were taken from the containers of 60 different manufactured foods. The lists averaged 17 words, with a range from 6 to 49. They were sorted by length into four comparable sets of 15. Within each set the lists were numbered 1 through 15. They were printed in 10-pt. Univers (Monotype series No. 689) with 1-pt. leading in two columns on a quarto sheet of paper 10 × 8 in. At the top of each list, next to its number, was the name of the food printed in capitals. The ingredients were

printed in lowercase, with capitals only for the first letter of each list and for the letter codes of vitamins, for example, vitamin B₁. The lists were set in unjustified lines which ranged in length from 25 to 35 letter spaces. Each list was reproduced photographically four times, in sizes to give about 10-, 7.5-, 6-, and 4-pt. letters (see Figure 1) with leading ranging from 1.0 to .4 pt. This insured that the words in a list did not change positions when the size of the lettering was changed.

One word in each list was selected as a target. For the first eight lists in each set, two of the target words were at the start of a line, two were at the end, and the remaining four were neither at the start of a line nor at the end. For the last seven lists (which most housewives did not reach), there were again two target words at the start and end of a line, but only three with intermediate positions. If a housewife searched each list only until she reached the target word, the number of words she would have needed to scan in the first eight lists of each set ranged from 76 to 87. For the remaining seven lists the number ranged from 74 to 94. A stenciled question sheet was prepared for each of the four sets of lists. Like the corresponding set of lists, it carried the list numbers and food names typed in capitals. Below each name was the single target word typed in lowercase.

There were also two practice sets, each comprising six lists, and two practice question sheets. The first practice set was stenciled like the question sheets; the second was printed in Univers, and reproduced four times, once with each of the four test sizes of letters.

Experimental Conditions

Two experiments were run under different levels of illumination corresponding, respectively, to supermarkets and domestic food closets. In the first there was a fairly uniform level of about 40 ftc. on the tops of the tables at which the housewives sat. The lighting came from four fluorescent lamp fittings hanging 9 ft. above the floor. In the second experiment the level of illumination was reduced to about 2 ftc., by wrapping layers of pale blue cotton sheeting around the fluorescent lamps. The illumination levels were measured using a photometer made by Salford Electrical Instruments Ltd.

In each experiment the four sizes of print and four sets of lists were arranged in a Latin-square design with four groups of housewives. The lists were always presented in the same order, thus confounding practice effects with list difficulty. There were altogether 32 housewives in the experiment with the bright lighting and 40 in the experiment with the dim lighting. Housewives were randomly allocated to conditions.

Procedure

The experiments were conducted on groups of about 20 housewives seated at tables. Each housewife was handed a cardboard file containing her

test passages and questions, and a black ball-point pen. The order of her experimental conditions differed from that of her immediate neighbors. The first practice set of lists was used to teach the procedure. The housewife had to read the target word on her question sheet, find the word in the corresponding practice list, and cross it out. She was asked to work through the list as quickly as possible. The size of print of the second practice set of lists was the same as that of the first experimental condition, to insure that the housewife was thoroughly familiar with the experiment before she started. The procedure of the second practice was identical with that used in the experiment.

Before each part of the experiment the housewife took a stenciled question sheet from the appropriate compartment of her file and placed it on her left if she was right-handed. When everyone was ready, she took the corresponding test sheet from her file, placed it on her right, and started searching for the target words. She was allowed 25 sec. only. No one ever crossed out more than 13 of the 15 target words in this time. The experiment and practice together took about 25 min.

Experimental Subjects

The 72 housewives were members of a panel maintained at the Applied Psychology Research Unit at Cambridge. Their ages ranged from 22 to 64 yr., with a median age of 46 yr. Seventy percent wore reading glasses for the experiment, and one more said that she ought to have done so. They were paid 7/6 per hour (about \$.90) for their services, plus traveling expenses.

RESULTS

Figure 1A gives the results of the very first experimental condition, before the housewives had tried any of the other sizes of lettering. Comparisons are subject to chance differences between the groups of housewives receiving each condition, but are free from transfer effects (Poulton & Freeman, 1966). In both the bright and the dim illumination the difference between the 4-pt. and 6-pt. type was reliable at the .05 level of significance on a two-tailed Mann-Whitney *U* test (Siegel, 1956, pp. 116-127). In the bright illumination the difference between 6-pt. and 10-pt. Univers was just reliable ($p < .05$) on a one-tailed test. In the dim illumination none of the differences between 6-, 7.5-, and 10-pt. type was reliable statistically ($p > .05$).

Figure 1B gives the results pooled over housewives and orders of conditions. The only comparisons subject to chance differences between the groups of housewives are between

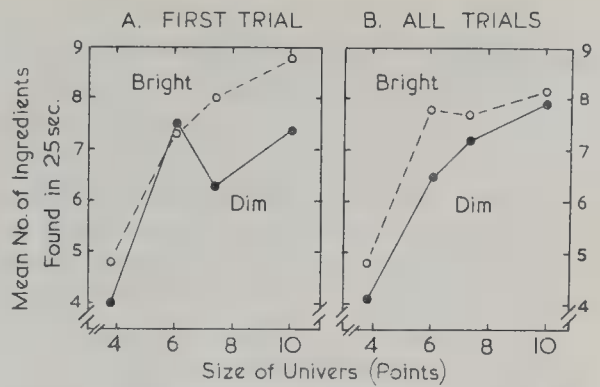


FIG. 1. Mean numbers of ingredients found in 25 sec. with different sizes of Univers lettering. (Unfilled points: about 40 ftc. on table top. Filled points: about 2 ftc. For the first trial, A, each unfilled point represents the mean performance of a separate group of 8 housewives, each filled point, the performance of 10 housewives. For all trials taken together, B, the broken line represents the performance of the same 32 housewives; the unbroken line is for 40 housewives.)

the two levels of illumination. Within each level of illumination the results for size of print are subject to unknown transfer effects from previous conditions with other sizes of lettering. An analysis of variance followed by Tukey's range test (Ryan, 1959, Appendix) was carried out separately on the data for each level of illumination. These statistical tests indicated that at both levels of illumination there was a highly ($p < .001$) reliable difference between the 4-pt. and 6-pt. type. In the bright illumination there was no statistically reliable difference between 6-, 7.5-, and 10-pt. Univers. But in the dim illumination the 10-pt. Univers was reliably better than any of the smaller sizes ($p < .05$ or better).

When presented with the small 4-pt. type in the dim light, 4 of the 40 housewives failed to locate any target words. All 4 were over 50 yr. old. Four other housewives located only one target word during the 25 sec. available. There were also 3 housewives who located only one target word in the bright light of the group of 32 housewives in this condition. The 7 housewives who located only one target word were all over 45 yr. old.

DISCUSSION

It is clear from Figure 1A and B that under certain conditions 10-pt. Univers may be

preferable to smaller sizes such as 7.5 or 6 pt. However, the main finding is the very large fall in the rate of locating target words when the size of the type was reduced from 6 to 4 pt. It suggests that 6-pt. Univers is about as small as food manufacturers should print the ingredients on containers if they are to be reasonably clear and legible to housewives of all ages.

The apparent size of lowercase print is determined principally by the x height (the height of the rounded parts of the letters, excluding the ascenders and descenders), not by the overall height or point size (Poulton, 1965, Figure 1). Univers is a modern sanserif face with an x height which is large in proportion to its face size. In order to find a conventional serified book typeface such as Modern Extended No. 1 (Monotype Series No. 7), Baskerville (No. 169), or Bembo (No. 270) with the same x height, it is necessary to add one-third to the point size (Poulton, 1965, Table 1). Thus, if 6-pt. Univers is taken as the minimal acceptable size for ingredients, this is equivalent to about 8-pt. Modern Extended, Baskerville, or Bembo.

The principal difference between the present results and those of Tinker and Paterson (1932) outlined previously must be due to the different age range of the experimental

subjects. Older people not only require a rather larger size of print for the optimal condition (Burt, 1959, p. 12), but also they cannot all read 4-pt. Univers in poor illumination. This has the same apparent size as 5.5-pt. normal book type.

REFERENCES

- BURT, C. *A psychological study of typography*. Cambridge University Press, 1959.
- LUCKIESH, M., & MOSS, F. K. *Reading as a visual task*. Princeton, N. J. Van Nostrand, 1942.
- POULTON, E. C. Letter differentiation and rate of comprehension in reading. *Journal of Applied Psychology*, 1965, **49**, 358-362.
- POULTON, E. C., & FREEMAN, P. R. Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin*, 1966, **66**, 1-8.
- RYAN, T. A. Multiple comparisons in psychological research. *Psychological Bulletin*, 1959, **56**, 26-47.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- TINKER, M. A. Validity of frequency of blinking as a criterion of readability. *Journal of Experimental Psychology*, 1946, **36**, 453-460.
- TINKER, M. A. *Legibility of print*. Ames: Iowa State University Press, 1963.
- TINKER, M. A., & PATERSON, D. G. Studies of typographical factors influencing speed of reading: IX. Reduction in size of newspaper print. *Journal of Applied Psychology*, 1932, **16**, 525-531.

(Received February 6, 1968)

KNOWLEDGE OF SCORE AND GOAL LEVEL AS DETERMINANTS OF WORK RATE¹

EDWIN A. LOCKE² AND JUDITH F. BRYAN

American Institutes for Research, Silver Spring, Maryland

Most previous studies of the motivational effects of knowledge of results have failed to control for differential goal setting by Ss in the different knowledge conditions. The present study attempted to separate the effects of knowledge qua knowledge from those of goal setting using a 2×2 factorial design. The task was simple addition. The factors were knowledge of (raw) score (KR) vs. no knowledge of (raw) score (No KR), and hard vs. easy goals. Scores in the KR condition were given in such a form that they could not be used to set goals. The hard- and easy-goal Ss, on the other hand, were informed only of their progress in relation to a standard set by E. It was found that the hard-goal Ss worked significantly faster than the easy-goal Ss, but the KR and No KR groups did not differ in performance.

The positive effects of knowledge of results (KR) on learning and performance are firmly established in the research literature (e.g., Ammons, 1956; Annett, 1961; Bilodeau & Bilodeau, 1961; Vroom, 1964). However, the question of how KR facilitates performance has not yet been answered.

The present paper is concerned only with the motivational function of KR. Thus, attention will be focused on those types of KR which have few or no cue or directive properties (i.e., KR which does not inform one of the nature and locus of one's errors or suggest how they might be corrected). Examples of purely or predominantly motivational KR would be knowledge of total score on a task summed over a number of trials or KR on simple psychomotor tasks which the subject (S) already knows how to perform. While it is true that KR of this type could be interpreted by S as a signal or cue to change his method of performing the task, it would not tell him what changes to make or how to go about correcting his errors. Such KR would have no directive function. (See Payne & Hauty, 1955, for a similar distinction between directive and motivational KR.)

¹ This research was supported by Grant No. MH 12103-01A1 from the National Institute of Mental Health.

² Now at the University of Maryland. Requests for reprints should be sent to Edwin A. Locke, Department of Psychology, University of Maryland, College Park, Maryland 20740.

A strong argument can be made for the thesis that KR affects performance indirectly: by influencing the nature of the goals the individual sets on the task. It can be observed by introspection that knowledge by itself does not have the power to initiate action. Man is necessarily selective in his use of information. The actions a man takes with respect to an item of knowledge depend upon the perceived significance of that knowledge to him. If an individual (in an experimental setting) appraises KR as signifying inadequate performance on a task, he will ordinarily set a goal to improve his subsequent performance. If he appraises KR as signifying adequate or superior performance, he will ordinarily set a goal to maintain or reduce his level of effort. And if the KR is appraised by him as irrelevant, he will take no action at all regarding it. Thus to predict performance, it is not sufficient to know that an individual was given KR; one must also know what he decided to do about it, that is, what goals he set in response to it.

If KR motivates performance through or by means of its effects on goal setting, this means that KR should have no effect on performance when differential goal setting is controlled.

In this context, it is relevant to ask whether effects attributed to (motivational) KR in previous studies could have been due to differential goal setting associated with the different KR conditions. An examination of the

literature on motivational KR indicates that the two variables have, in fact, been frequently confounded. (For a detailed review of these studies see Locke, Cartledge, & Koeppel, 1968.)

Book and Norvell (1922), Crawley (1926), and Mackworth (1950), for example, assigned their KR Ss specific goals to aim for, while No KR Ss were typically told to "do their best." (Previous studies have found that specific hard goals lead to a higher performance level than a goal of "do your best"; Locke & Bryan, 1967.)

In other studies (Adams & Humes, 1963; Church & Camp, 1965; McCormack, Binding, & Chylinski, 1962; McCormack, Binding, & McElheran, 1963; Payne & Hauty, 1955) goals were not explicitly assigned to Ss, but KR was always expressed in relation to a standard. For example, KR Ss would be told whether or not they had surpassed their immediately previous score on the task. When an experimental *S* is informed that he has failed to surpass a standard set by the experimenter (*E*), it is highly likely that he will "try harder" on subsequent trials. Similarly he might be expected to relax somewhat if told that his performance had exceeded expectations.

There have been a number of studies of (motivational) KR in which there was no obvious manipulation of goal setting by *E* (Arps, 1920; Brown, 1932; Johanson, 1922; Mace, 1935; Manzer, 1935), but the possibility that Ss in the KR and No KR conditions set different goals spontaneously cannot be ruled out. Mace, in fact, explained his results by arguing that the KR Ss set themselves harder implicit goals on the task than the No KR Ss.

The potential importance of controlling for differential goal setting was indirectly emphasized in a recent study by Chapanis (1964). He "hired" his Ss as employees rather than as "experimental" Ss and ran them individually for an hour a day for 24 days. The task was punching digits onto a paper tape. Chapanis found no differences in the output of the KR and No KR groups. Two characteristics of this study that may have lessened the motivation of the KR Ss to set specific goals were (a) the probable absence of the "demand

characteristics" (Orne, 1962) which are inherent in most experimental situations, such as the implicit demand to "improve" one's performance, and (b) the absence of the opportunity for inter-*S* competition which is present in many studies of this type (e.g., Gibbs & Brown, 1955).

Both the foregoing discussion and previous findings in this area point to the desirability of experimentally separating the effects of KR per se from those of goal setting. In a previous study using a 2×2 design, Locke (1967a) gave half his Ss knowledge of their actual scores on an addition task, while half received no KR. The opportunity for differential goal setting by these groups was reduced by using trials of alternating (10 and 15 min.) lengths, so that scores on consecutive trials were not comparable. Half the Ss were told to "do their best" on the task, while half were assigned specific hard goals to reach. The goals were indicated by colored cards placed in Ss' box of problem cards designating how far they were supposed to get on that trial. No effect of KR condition on performance was found in this study, whereas there was a significant goal effect in favor of the hard-goal Ss.

The present study was also designed to separate KR and goal-setting effects, but incorporated a number of methodological improvements over the previous study by Locke (1967a). In that study the No KR Ss could have gotten some idea of their progress, as the box of problems cards was clearly visible in front of them. Further, the alternating trial lengths did not prevent Ss from comparing their scores on every other trial and setting personal goals. Finally the hard-goal-No KR Ss could have obtained knowledge of their scores by observing where the goal card was placed in the box. The addition task was structured in the present study so that these problems were avoided.

A further problem with the previous study concerned the "do best" goal. Although this goal was used purposely in order to duplicate the goals that had been used by earlier investigators (Book & Norvell, 1922; Crawley, 1926; Locke & Bryan, 1966), there was no way of specifying a priori just what level of motivation such a goal would induce. In the

present study it was decided to use specific hard goals and specific easy goals rather than hard and "do best" goals. Numerous previous studies have indicated that hard goals consistently produce a higher level of performance than easy goals (see Locke, 1968, for a summary), thus the same prediction was made for the present study. No effect of KR per se on performance was anticipated.

It should be mentioned that although it is possible to give KR in such a form that goal setting is completely ruled out, it is not possible to set Ss' goals without giving them some knowledge, either of the goals themselves and/or of their progress in relation to these goals. In the present study the terms "hard goal" and "easy goal" will be used to describe groups given knowledge of their progress in relation to goals set by *E*, while the terms KR and No KR will be used to refer to groups given or not given knowledge of their actual (raw) scores on the task.

METHOD

Subjects

The Ss were 23 male and 17 female University of Maryland volunteers who were paid for participation. The Ss were run individually.

Task

The task was simple addition; each problem consisted of three two-digit numbers. The problems were typed on a roll of paper which was wound on a spool. The spool was placed in a box with a transparent window. The *S* advanced one problem at a time to the window by turning a knob. Answers were written on separate answer sheets which contained spaces for 60–90 answers. (The number and arrangement of the spaces on the page varied from sheet to sheet.) Problems on the spool were matched with answer spaces by means of random numbers which identified each problem and its corresponding answer space. These numbers told *S* where to write his answers, but could not reveal how many problems he had done. (The *S* did not have to search the whole page for the right space since the answer sheets were set up so that *S* worked across the page line by line, as in reading a book.) When an answer sheet was completed, *S* began immediately on the next sheet. At the end of the trial all sheets for that trial were handed in to *E* in an adjoining observation room through a hole in the wall.

There were five trials, each of a different length, ranging from 8 min. 45 sec. to 15 min. 30 sec. ($M = 12$ min.). Trials were separated by 3-min. rest periods. The Ss were interrupted at specific intervals during each trial and given feedback as described

below. There were a total of 30 interruptions during the five trials, the average interval between interruptions being 2 min. (range: 50 sec. to 3 min. 25 sec.). The Ss knew that the interruption periods and trial times all differed in length, but did not know what the lengths were.

Design and Procedure

The design was a 2×2 factorial model with 10 Ss per cell. The variables were KR versus No KR and hard versus easy goals. To assign Ss to cells, the 40 Ss were first ranked in order of ability as determined by a 3-min. addition pretest. The four Ss of lowest ability were then assigned at random to one of the four cells, the next four Ss were treated identically, and so on until all Ss had been assigned to a cell. This had the effect of equalizing the cell ability means while retaining random selection.

No knowledge of score (No KR). Half the Ss were not given their scores at any time during the experiments.

Knowledge of score (KR). Half the Ss were told the cumulative number of problems they had attempted on that trial at each interruption point. These interruptions were always made after *S* had just completed a problem and was going to the next one, never when *S* was actually working on a problem. At the end of each trial, *S* was told how many problems he had gotten correct on that trial. (This could not be done during the trial as *E* did not have the answer sheets.)

Easy goal. Half the Ss were assigned easy output goals, in terms of number correct, to work for on each trial. To reach these goals, *S* had to work at 67% of his rate (problems correct per minute) on the 3-min. practice trial. The *S* was told that the assigned rate was slower than his practice trial, but was not given the actual percentage figure.

The Ss were not told their actual numerical goals, but were informed of their progress in relation to their goal by means of a set of four lights displayed on a console. The illumination of the lights was controlled by *E*. When a 1-in. (diameter) red light was illuminated, this indicated that *S* was more than three problems behind the pace required to reach his goal for that trial; the illumination of a $\frac{1}{2}$ -in. red light indicated that *S* was one to three problems behind his assigned pace; the illumination of a $\frac{1}{4}$ -in. green light indicated *S* was one to three problems ahead of his assigned pace; and the illumination of a 1-in. green light indicated that *S* was more than three problems ahead of his assigned pace for that trial. The *E* turned on the appropriate light at each interruption point to indicate *S*'s rate in relation to his goal at that point in the trial.

The Ss were informed in advance as to what the different lights signified and were instructed to speed up in response to the red lights. They were told that the lights indicated their performance relative to their goal only in terms of number attempted. The *E* kept track of *S*'s number of problems attempted by observing his progress through a one-way mirror. At the end of each trial (after *E* had corrected the

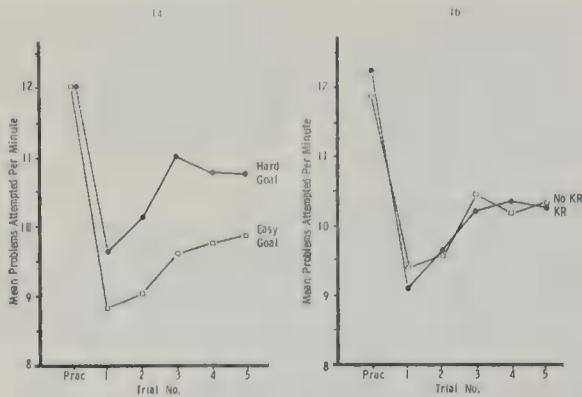


FIG. 1. Mean problems attempted per minute as a function of amount of knowledge of score and goal level.

answer sheets), *S* was told whether or not he had beaten his goal for that trial in terms of number correct.

Hard goal. Half the *Ss* were assigned hard output goals to reach on each trial. To reach his goal, a hard-goal *S* had to work at the same pace he had worked at during the 3-min. practice trial. The *Ss* were told that they had to work at 100% of their practice-trial pace in order to reach their goals. Feedback was given to these *Ss* with lights as with the easy-goal *Ss*.

To summarize, each *S* was interrupted 30 times during the five trials. The *S* was told his cumulative number of problems attempted (if in the KR condition) and was informed (by means of lights) of his work pace in relation to his assigned goal. At the end of each trial *S* was told his number correct (if appropriate) and whether or not he had beaten his goal on that trial.

The design of the task insured that KR and goal setting were separated. The KR *Ss* could not use their knowledge to set personal goals, and the No KR *Ss* could not use the goal feedback to get knowledge of their actual scores. The goal *Ss*, of course, had some knowledge, but it was expressed only as a relation between their performance and a standard set by *E*.

The *E* remained in an observation room during the experiment, but kept track of *S*'s progress through a one-way mirror (of which *S* was made aware). Communication was possible through an intercom system.

RESULTS

Three different performance criteria were used: (a) deterioration in number attempted, defined as the difference between mean number of problems attempted per minute on the practice trial and the mean number attempted per minute in the experimental period (the deterioration in rate was due to the experimental trials being longer than the

practice trial), (b) deterioration in number correct, same as above, but in terms of number correct, and (c) percentage of errors, defined as the total number of problems wrong/total number of problems attempted during the experimental period.

The number attempted curves for the hard- and easy-goal *Ss* are shown in Figure 1a, and the corresponding curves for the KR and No KR *Ss* are shown in Figure 1b. It is evident that the hard-goal *Ss* maintained a faster work pace than the easy-goal *Ss*, but the KR and No KR *Ss* worked at virtually the same pace.

The data were analyzed using a standard 2×2 factorial analysis of variance. The goal effect was significant for the number attempted criterion, $F(1, 36) = 5.50$, $p < .05$, and for the percentage of errors criterion, $F(1, 36) = 7.67$, $p < .01$, the hard-goal *Ss* attempting significantly more problems and making significantly more errors than the easy-goal *Ss*. The difference for the number correct criterion also favored the hard-goal *Ss*, but did not reach significance, $F(1, 36) = 2.01$, $p > .05$.

There was no significant KR effect for any of the criteria (all $Fs < 1.02$), nor were there any significant interaction effects (all $Fs < 1$).

The work rate (number attempted) of the two goal groups as a function of the length of the work period between interruptions was also compared. Dividing the intervals into those under 1 min. ($n = 8$), those between 1 and 2 min. ($n = 14$), and those over 2 min. ($n = 13$), a two-factor repeated-measures analysis of variance was performed. There was a significant interaction between goal group and interval length, $F(2, 76) = 3.92$ ($p < .05$). The hard-goal group worked at a faster rate during the longer intervals than during the shorter intervals, while the opposite was true for the easy-goal group.

The empirical difference in goal difficulty between the hard and easy goals can be seen by comparing the distribution frequencies of the various feedback lights, shown in Table 1. The hard-goal *Ss* had large or small green light feedback (indicating they were ahead of their assigned pace in number attempted) only 37% of the time, whereas the easy-goal

Ss had green light feedback 96% of the time. The chi-square value based on the frequency distribution shown in Table 1 is 536.38 ($p < .001$, $df = 3$). In terms of number of problems correct for the trials as a whole, the hard-goal Ss beat their goals on 10% of the trials, while the easy-goal Ss beat their goals on 86% of the trials.

The KR and No KR Ss did not differ significantly in the distribution of feedback lights encountered nor in the percentage of success in reaching their trial (number correct) goals.

Interestingly, when the actual mean work rates (in relation to practice-trial rate) were compared after red versus green light feedback, no differences emerged within either goal group. The hard-goal Ss worked at a relatively faster rate than the easy-goal Ss after *both* red and green light feedback. The t for the mean difference in rate after green light feedback was 2.88 ($p < .01$). The corresponding t for red light feedback was only 1.64 ($p < .05$), but this lower value may have been due to the very small number of easy-goal Ss ($n = 7$) who experienced red light feedback. The absolute mean difference in work rate between the two goal groups was the same for both green and red light feedback.

These findings indicate that the faster work rate of the hard-goal Ss cannot be attributed simply to the greater amount of red light feedback they received as compared with the easy-goal Ss. The hard-goal Ss adopted a generally faster work pace in response to all the lights. This is not surprising in view of the fact that the hard-goal Ss had to work 33% faster than the easy-goal Ss to get the same light (the large red light excluded) to come on; that is, the standards regulating the light feedback were higher for one group than the other. If one assumes that all Ss were trying to get the green lights to come on, then the fact that one group had to work faster to do this than the other would explain their faster overall rate.

DISCUSSION

The finding that KR had no effect on performance when goal setting was controlled supports the results of Locke's (1967a) study in which the two variables were less defini-

TABLE 1
FREQUENCY OF FEEDBACK LIGHTS AS A FUNCTION OF GOAL CONDITION

| Goal | Feedback light | | | | Total |
|------|----------------|-----------|-------------|-------------|-------|
| | Large red | Small red | Small green | Large green | |
| Hard | 250 | 129 | 121 | 100 | 600 |
| Easy | 5 | 22 | 115 | 458 | 600 |

Note.—There were 20 Ss in each goal condition, and each S had feedback lights on 30 different occasions. For an explanation of the meaning of each light, see text.

tively separated. The positive relationship obtained between goal difficulty and performance level also supports the results of a number of previous studies (Day & Kaur, 1965; Locke, 1966, 1967b, 1968; Stedry, 1960).

The fact that the hard-goal Ss surpassed the easy-goal Ss in number attempted but not in number correct is not surprising if one considers that KR during the trials was given in terms of rate (i.e., number attempted) rather than in terms of number correct. The greater number of errors achieved by the hard-goal Ss can be attributed to their trying to speed up at the expense of accuracy.

No definite explanation can be given for interaction between pace and length of work interval for the two goal groups, but some hypotheses may be suggested. The slower rate of the easy-goal Ss during the longer intervals could be the result of a relaxation of effort over time on their part; such relaxation would be encouraged by their knowledge that they were (usually) ahead of their required pace. In contrast, the hard-goal Ss, who were typically behind their required pace, may have been overly tense during the short trials, but may have relaxed enough during the longer trials to increase their rate.

The present findings indicate that the results of previous studies in which KR and goal setting were simultaneously manipulated could be attributed solely to goal-setting effects. However, it remains to explain the results of KR studies in which goal setting was not explicitly manipulated (e.g., Arps, 1920; Brown, 1932; Johanson, 1922; Mace, 1935; Manzer, 1935). Extrapolating from the present findings, it may be hypothesized that

the KR groups in these previous studies independently set themselves harder goals than did the No KR groups. A recent study by Locke and Bryan (1968) supports such an interpretation. One group of Ss was given KR on each of 16 5-min. trials of a computation task, while another group did not receive KR. No goal-setting instructions were given. The KR group performed better than the No KR group on the last eight trials, but only because the KR Ss spontaneously set harder goals on these trials than the No KR Ss. When goal differences were controlled by partialing, the KR effect was vitiated. This finding supports the view that when KR does facilitate performance it does so through its effects on goal setting.

This finding is perhaps not very surprising. It is difficult to imagine how KR could motivate performance unless it were given in such a form that the individual could appraise it in some manner. But appraisal requires a standard against which performance can be judged. Typically an S uses his own previous performance or the performance of another S (or, in the case of some experiments, standards supplied by E). Such standards, of course, were not provided by the KR given in the present study. When questioned about the use they made of the knowledge of their scores, the KR Ss agreed unanimously that it was of no use to them whatever, and that they had therefore ignored it.

These findings imply further that KR should affect performance most when goal setting is most facilitated (assuming S desires to improve). Clearly goal setting is easier when all the trials are of the same length than when they are all of different lengths; in the former case S has some standard by which he can judge and evaluate his performance and set new goals, whereas in the latter case he does not. Further, KR which leads to the setting of hard goals should be more effective than KR which leads to the setting of easy goals. The setting of hard goals could be facilitated by giving KR in relation to hard standards (e.g., S's best previous score) and the setting of easy goals by giving KR in relation to easy standards (e.g., S's worst previous score).

The finding that KR has no effect on per-

formance independent of goal setting is congruent with results obtained in studies of other incentives, including money (Locke, Bryan, & Kendall, 1968), time limits (Bryan & Locke, 1967), and "verbal reinforcement" (Dulany, 1968; Holmes, 1967; Spielberger, Bernstein, & Ratliff, 1966). These incentives were found to affect performance only if and to the extent that they affected Ss' conscious goals and intentions. The evidence indicates that more attention should be paid to the goals and intentions which S develops in response to the incentives E provides. It would also be of interest to study how the form in which incentives are given affects the goals which Ss develop on tasks.

REFERENCES

- ADAMS, J. A., & HUMES, J. A. Monitoring of complex displays: IV. Training for vigilance. *Human Factors*, 1963, 5, 147-153.
- AMMONS, R. B. Effects of knowledge of performance: A survey and tentative theoretical formulation. *Journal of General Psychology*, 1956, 54, 279-299.
- ANNETT, J. The role of knowledge of results in learning: A survey. Technical Report 342-3, 1961, United States Naval Training Device Center.
- ARPS, G. F. Work with knowledge of results versus work without knowledge of results. *Psychological Review Monograph Supplement*, 1920, 28(3, Whole No. 125).
- BILODEAU, E. A., & BILODEAU, I. McD. Motor-skills learning. In *Annual review of psychology*. Palo Alto, Calif.: Annual Reviews, 1961.
- BOOK, W. F., & NORVELL, L. The will to learn: An experimental study of incentives in learning. *Pedagogical Seminary*, 1922, 29, 305-362.
- BROWN, F. J. Knowledge of results as an incentive in school room practice. *Journal of Educational Psychology*, 1932, 23, 532-552.
- BRYAN, J. F., & LOCKE, E. A. Parkinson's law as a goal-setting phenomenon. *Organizational Behavior and Human Performance*, 1967, 2, 258-275.
- CHAPANIS, A. Knowledge of performance as an incentive in repetitive monotonous tasks. *Journal of Applied Psychology*, 1964, 48, 263-267.
- CHURCH, R. M., & CAMP, D. S. Change in reaction time as a function of knowledge of results. *American Journal of Psychology*, 1965, 78, 102-106.
- CRAWLEY, S. L. An experimental investigation of recovery from work. *Archives of Psychology of New York*, 1926, No. 85.
- DEY, M. K., & KAUR, G. Facilitation of performance by experimentally induced ego motivation. *Journal of General Psychology*, 1965, 73, 237-247.
- DULANY, D. E., JR. Awareness, rules and propositional control: A confrontation with S-R behavior

- theory. In D. Horton & T. Dixon (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.
- GIBBS, C. B., & BROWN, I. D. Increased production from the information incentive in a repetitive task. (No. 230) Cambridge, England: Medical Research Council, Applied Psychology Research Unit, 1955.
- HOLMES, D. S. Verbal conditioning or problem solving and cooperation? *Journal of Experimental Research in Personality*, 1967, 2, 289-294.
- JOHANSON, A. N. Influence of incentive and punishment on reaction time. *Archives of Psychology*, 1922, 8, No. 54.
- LOCKE, E. A. The relationship of intentions to level of performance. *Journal of Applied Psychology*, 1966, 50, 60-66.
- LOCKE, E. A. The motivational effects of knowledge of results: Knowledge or goal setting? *Journal of Applied Psychology*, 1967, 51, 324-329. (a)
- LOCKE, E. A. Relationship of goal level to performance level. *Psychological Reports*, 1967, 20, 1068. (b)
- LOCKE, E. A. Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, 1968, 3, 157-189.
- LOCKE, E. A., & BRYAN, J. F. Cognitive aspects of psychomotor performance: The effects of performance goals on level of performance. *Journal of Applied Psychology*, 1966, 50, 286-291.
- LOCKE, E. A., & BRYAN, J. F. Performance goals as determinants of level of performance and boredom. *Journal of Applied Psychology*, 1967, 51, 120-130.
- LOCKE, E. A., & BRYAN, J. F. Goal-setting as a determinant of the effect of knowledge of score on performance. *American Journal of Psychology*, 1968, in press.
- LOCKE, E. A., BRYAN, J. F., & KENDALL, L. M. Goals and intentions as mediators of the effects of monetary incentives on behavior. *Journal of Applied Psychology*, 1968, 52, 104-121.
- LOCKE, E. A., CARTLEDGE, N., & KOEPPPEL, J. The motivational effects of knowledge of results: A goal-setting phenomenon? *Psychological Bulletin*, 1968, 70, 474-485.
- MACE, C. A. Incentives: Some experimental studies. Industrial Health Research Board (Great Britain), 1935, Report No. 72.
- MACKWORTH, N. H. Researches on the measurement of human performance. (Special rep. series, No. 268) Cambridge, England: Medical Research Council, Applied Psychology Unit, 1950.
- MANZER, C. W. The effect of knowledge of output on muscular work. *Journal of Experimental Psychology*, 1935, 18, 80-90.
- MCCORMACK, P. D., BINDING, F. R. S., & CHYLYNSKY, J. Effects on reaction-time of knowledge of results of performance. *Perceptual and Motor Skills*, 1962, 14, 367-372.
- MCCORMACK, P. D., BINDING, F. R. S., & McELHERAN, W. G. Effects on reaction-time of partial knowledge of results of performance. *Perceptual and Motor Skills*, 1963, 17, 279-281.
- ORNE, M. T. On the social psychology of the psychological experiment with particular reference to demand characteristics. *American Psychologist*, 1962, 17, 776-783.
- PAYNE, R. B., & HAUTY, G. T. Effect of psychological feedback upon work decrement. *Journal of Experimental Psychology*, 1955, 50, 343-351.
- SPIELBERGER, C. D., BERNSTEIN, I. H., & RATLIFF, R. G. Information and incentive value of the reinforcing stimulus in verbal conditioning. *Journal of Experimental Psychology*, 1966, 71, 26-31.
- STEDRY, A. C. *Budget control and cost behavior*. Englewood Cliffs, N. J.: Prentice-Hall, 1960.
- VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

(Received February 7, 1968)

OFF-QUADRANT COMMENT

NORMAN M. ABRAHAMS¹

United States Naval Personnel Research Activity, San Diego, California

This discussion points out a serious methodological problem of quadrant analysis (QA), a method for identifying correlates of individual differences in predictability. QA involves division of the predictor-criterion scatterplot into four cells by cutting at the predictor and criterion medians. For cases below the *predictor* median, and separately, for cases above the predictor median, a comparison is made between the test responses and scores of those above the *criterion* median with those below. The distinguishing responses and scores are combined into moderator variables, one for each predictor group. To the extent of the relationship between the predictor and criterion, cases above and below the criterion median within each predictor category are *not equivalent* on predictor scores. Because of these differences, items or scores found to differentiate between comparison groups may merely reflect the predictor composite differences. Moderators developed through QA may therefore be predictive only of the composite predictor itself.

The apparent ceiling in predictive validity, well documented by Ghiselli (1955), stimulated a search for new approaches to improve behavioral prediction. Primarily, this need was met with pleas for abandonment of the simple one-to-one predictor-criterion correlational model (see, e.g., Dunnette, 1963; Guetzkow & Forehand, 1961; Primoff, 1955). Perhaps the most intriguing, promising, and convenient weapon in the struggle toward improving prediction was the concept of the moderator variable (Saunders, 1956). While several approaches were offered (Banas, 1964; French, 1961; Lykken, 1956; Marks, undated; Rimland, 1960; Toops, 1959), Ghiselli's (1956, 1960) had the greatest appeal due to its directness of application and empirical support. Briefly, this technique analyzes data to identify items or variables that correlate with absolute differences, *Ds*, between standard scores on the predictor and standard scores on the criterion. The appearance of this method resulted in a widespread resurgence of interest in moderating simple predictions. All at once it seemed that everyone analyzed his data for moderator variables, and it ap-

peared from published reports that they were being found—but only as often as chance permitted. These discouraging results led many researchers to throw in the tear-stained moderator towel and search elsewhere for enhanced prediction.

Unlike many less Spartan researchers, however, Hobert (1965; Hobert & Dunnette, 1967) remained undaunted when the *D*-score strategy failed to produce positive results. Instead, Hobert developed an alternative procedure, named quadrant analysis, for discovering correlates of individual differences in predictability. Using managerial performance as the criterion and a composite score as the predictor, he partitioned the predictor-criterion scatterplot as shown in Figure 1 into four quadrants by first cutting at the overall predictor median and then on the criterion score median. Next, since the relationship was positive ($r = .70$), he labeled the upper-left and lower-right quadrants as underpredicted and overpredicted, respectively. Analogously, the upper-right and lower-left quadrants were labeled as high hits and low hits, respectively.

In attempting to uncover a moderator variable that would discriminate between the "low hits" and the "underpredicted" cases, Hobert contrasted the two groups' item responses and test scores on predictors in the original predictor set. Similarly, the high hits and overpredicted managers were compared on the same responses and scores. For each com-

¹ The opinions expressed are those of the author and do not necessarily reflect those of the Navy Department. The author is indebted to Marvin D. Dunnette, University of Minnesota, for his helpful suggestions on an earlier draft of this article. Requests for reprints should be sent to the author, United States Naval Research Activity, San Diego, California 92152.

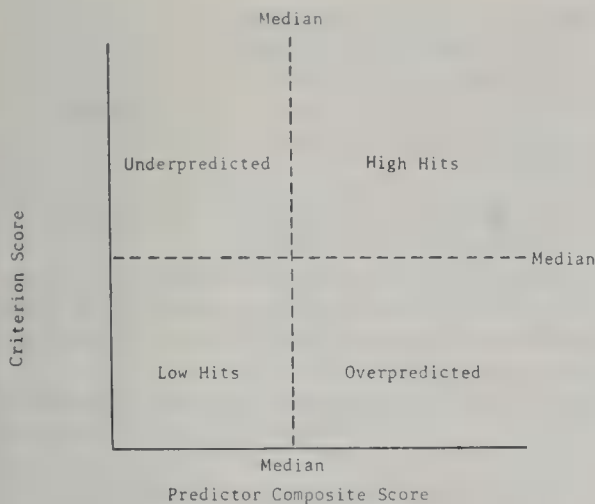


FIG. 1. Classification of cases by quadrant analysis.

parison, differences between the groups were combined into a "moderator" that hopefully would provide enhanced predictions on future subjects. This procedure produced two moderators, one for low- and one for high-predictor cases. The moderators were cross-validated on hold-out samples and appeared to improve prediction.

In comparison with other approaches, Hobert cited several potential advantages for identifying moderator variables in this manner. First, under- and overpredicted cases are not considered jointly as they are in the *D*-score approach. One moderator for overpredicted and one for underpredicted cases are constructed. Since failure is generally more predictable than success, it may be expected that differential validities would occur for high- and low-predictor cases, and quadrant analysis can take account of such differences. Thus, the technique does not involve dilution of possible differences between high- and low-predictor errors, unavoidable with most other approaches. Finally, from a practical standpoint, the four subgroups in quadrant analysis are obtained easily, requiring no *D*-score type computations.

This procedure, quadrant analysis, while intuitively logical and appealing, incorporates a serious shortcoming that could easily affect the results. This shortcoming may best be understood by referring to the scatterplot in Figure 2.

The shortcoming of this procedure is that mean differences on the predictor composite

between comparison groups are ignored. For the low-predictor groups, that is, low hit and underpredicted, predictor means have been estimated by inspection and are indicated in Figure 2. Similarly, for high hits and overpredicted groups, approximate mean predictor scores are indicated.

Obviously, there are sizable differences on the predictor-score means which are ignored when underpredicted and overpredicted groups are contrasted with high and low hits. Differences found in item analyses may merely reflect these predictor differences. In other words, the so-called moderator variable may only be predictive of differences on the predictor variable itself. At the very best, it appears that the components of the moderator variable would reflect a combination of differences, that is, actual differences between hits and misses and spurious differences due to predictor-score differences.

Thus, quadrant analysis virtually insures finding differences between hits and miss-predicted cases. When items and variables involved in the original predictor set are searched for possible inclusion in the moderator scale, as in the Hobert study, it is further assured that differences will be found. And, the differences found will most likely be those already receiving the greatest weight in the original predictor composite. Since the variables showing such differences will be so

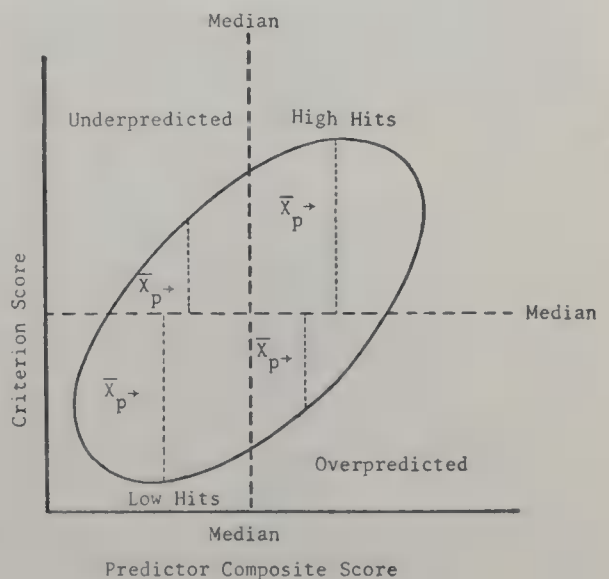


FIG. 2. Estimated predictor-score means for each predictor-criterion category.

highly related to the predictor, it is doubtful that such a moderator should be expected to have value beyond that already shown in the predictor-criterion relationship. The net effect of using data in the original predictor set as a moderator is equivalent to changing the weighting used to obtain the composite score. Hobert's and Dunnette's (1967) suggestion that moderators using items outside the original predictor set may be more efficient should be taken seriously in subsequent applications of the quadrant approach. However, this suggestion does *not* remove the problem imposed by the built-in composite-predictor score differences between comparison groups. Perhaps this difficulty could be overcome by matching comparison groups on the predictor-composite score.

The problem of improving prediction with moderator variables is an interesting and important one since in practical situations predictive accuracy is rarely high. Hobert has clarified some of the disadvantages of existing moderator approaches and suggested an interesting alternative, but a solution to the methodological limitations of quadrant analysis is yet to be found.

REFERENCES

- BANAS, P. M. An investigation of transsituational moderators. Unpublished doctoral dissertation, University of Minnesota, 1964.
- DUNNETTE, M. D. A modified model for test validation and selection research. *Journal of Applied Psychology*, 1963, **47**, 317-323.
- FRENCH, J. W. *A machine search for moderator variables in massive data*. Princeton, N. J.: Educational Testing Service, 1961.
- GHISELLI, E. E. *The measurement of occupational aptitude*. Berkeley: University of California Press, 1955.
- GHISELLI, E. E. Differentiation of individuals in terms of their predictability. *Journal of Applied Psychology*, 1956, **40**, 374-377.
- GHISELLI, E. E. Differentiation of tests in terms of the accuracy with which they predict for a given individual. *Educational and Psychological Measurement*, 1960, **20**, 675-684.
- GUETZKOW, H., & FOREHAND, G. A. A research strategy for partial knowledge useful in the selection of executives. In R. Taguiri (Ed.), *Research needs in executive selection*. Boston: Harvard Graduate School of Business Administration, 1961.
- HOBERT, R., & DUNNETTE, M. D. Development of moderator variables to enhance the prediction of managerial effectiveness. *Journal of Applied Psychology*, 1967, **51**, 50-64.
- HOBERT, R. D. Moderating effects in the prediction of managerial success from psychological test scores and biographical factors. Unpublished doctoral dissertation, University of Minnesota, 1965.
- LYKKEN, D. T. A method of actuarial pattern analysis. *Psychological Bulletin*, 1956, **53**, 102-107.
- MARKS, M. R. *Derivation of moderator variables: Theoretically and from data*. Rochester: University of Rochester, undated.
- PRIMOFF, E. S. *Test selection by job analysis*. Washington, D. C.: United States Civil Service Commission, Test Development Section, 1955.
- RIMLAND, B. Multidimensional scatterplotting: A graphic approach to profile analysis. *Journal of Applied Psychology*, 1960, **44**, 404-406.
- SAUNDERS, D. R. Moderator variables in prediction. *Educational and Psychological Measurement*, 1956, **16**, 209-222.
- TOOPS, H. A. A research utopia in industrial psychology. *Personnel Psychology*, 1959, **12**, 189-225.

(Received February 14, 1968)

MODERATION OF A MODERATOR TECHNIQUE

QUINN McNEMAR ¹

University of Texas at Austin

A modification of a method proposed by Hobert and Dunnette (1967) for finding moderator variables is developed and then illustrated by way of an idealized set of data. It is shown that in terms of increased validity coefficients, decreased percentage of overlap, and increased hit rate the modification, which is simple and very easy to use, leads to marked improvement over the original Hobert and Dunnette method. Then the seriously misleading nature of this search for moderators, shared alike by the original and modified versions, is pointed out.

The paper by Hobert and Dunnette (1967) reports the dissertation efforts of Hobert to use quadrant analysis, an operational method for finding moderator variables that emerged from Dunnette's (1963, 1966) modification of a model proposed by Guetzkow and Forehand (1961) for test validation and/or prediction. There is no need to describe the model here, but a brief statement of quadrant analysis is required in order to understand a paradox that can arise from its usage.

Imagine the bivariate scattergram for a criterion, Y, and a predictor, X. By cutting at, say, the medians the resulting four-fold table permits a classification of individuals into "high hits" (high predictor, high criterion, in upper-right quadrant), "low hits" (low predictor, low criterion, in lower-left quadrant), "overpredicted" (high predictor, low criterion, in lower-right quadrant), and "underpredicted" (low predictor, high criterion, in upper-left quadrant). Now consider the two low-predictor groups, the low hits and the underpredicted. Obviously, they will differ rather markedly on the criterion, or Y, variable, whereas (it is asserted that) the two groups "have common predictor," or X, scores. (By symmetry, similar characterization would hold for the two high-predictor groups.) If items or variables could be found that differentiate the low hits and underpredicted groups, one would have the where-withal for identifying those that would be underpredicted by X. Thus a "moderator" variable would have been found.

By using this quadrant analysis, Hobert and Dunnette (1967) did find scales that when used as a moderator test eliminated 25% as unpredictable and raised the validity from .65 to .73. This apparent success in the chase for moderator variables does, however, need further scrutiny before too much effort is spent thereon.

In order to examine the situation, a two-way symmetric frequency table (Figure 1) was generated for $N = 1,000$, by use of Pearson's tables, for a normal bivariate correlation of .70 which was chosen because of its nearness to the validities in the Hobert and Dunnette study. Actually, the calculated r for Figure 1 is .687, which deviates from .70 because of rounding of the frequencies, ignoring cases that deviate more than 2.7 sigmas, and the grouping error (Sheppard's



FIG. 1. Theoretical normal bivariate scatter for r of near .70: Y = criterion, X = predictor.

¹ Requests for reprints should be sent to the author, Department of Psychology, University of Texas, Mezes Hall 211, Austin, Texas 78712.

correction raises the r to .692). For convenience, both X and Y are represented by scores ranging from 1 to 18, $M_s = 9.5$, $SD_s = 3.296$.

It is relevant to report some statistics (actually parameters since one is dealing with a theoretic bivariate distribution). For all of the low-predictor cases (left half of scatter): M for $X = 6.84$, M for $Y = 7.68$, the SD_s are 1.95 and 2.75, respectively. For the under-predicted cases (upper left), M for $X = 7.87$, M for $Y = 11.13$; for the low hits (lower left), M for $X = M$ for $Y = 6.49$. The large differences between the Y means would certainly be anticipated, but these two subgroups did not achieve exactly "similar predictor" (X) scores. The difference in means, $7.87-6.49$, is sufficiently large that *if* based on samples even as small as 10 and 30 (proportional to quadrant frequencies in Figure 1) significance would be claimed at the .01 level, two-tailed test.

From the fact that the low hits and under-predicted subgroups differ not only on Y , the criterion, but also on the predictor (X), it follows that any items or variables that differentiate between these two groups may do so for either of two obvious reasons. This would seem to be advantageous in that discriminating items (and/or variables) for use in a moderator test need not be uncorrelated with the predictor variable, X . (If it were known that this type of search did lead to a moderator that is uncorrelated with X , it would of course be good strategy to combine it with X by multiple regression in order to enhance the prediction of Y .)

Now Hobert and Dunnette searched for, and found, discriminating items (and scales) right in the pool of items (and scales) making up the predictor battery. Unweighted scoring of these items (and scales) produced an "item" and a "scale" moderator test. Both yielded correlations with the criterion, but the scale moderator held up better under cross-validation. With a test that does "discriminate between two subgroups of individuals both of whom attained similar scores on a test battery, but different criterion scores [p. 57]," their next step was to use it to identify those for whom predictions based

on the test battery will be incorrect. Getting rid of these would leave those that are more predictable. The method of identification will be given below.

This reuse of items and scales from the original battery, with no requirement that the moderator test thus built be uncorrelated with the predictor (i.e., the original battery), led the present author to hypothesize that the predictor itself would prove to be a moderator. If so, one could forget the search for items for the moderator test, and assuming cross-validation for the predictor one could also forget about cross-validation for the moderator (since there is no capitalization on chance at this stage).

Refer again to Figure 1 in which it is readily seen that some correlation must hold for the 500 cases in the left half of the scattergram. Straightforward calculation leads to an r of .49 (a value which may be regarded as a parameter—not subject to sampling error). Within the half, the regression for criterion on the "moderator" is exactly linear, and the fact that the other regression is curvilinear is of no relevance. Obviously the "moderator" discriminates between the under-predicted and the low-hits group, hence it can be used to identify some of those that would be unpredictable. (The low hits, as well as the high hits, are regarded as the predictables.)

The procedure used by Hobert and Dunnette is next followed for ascertaining a cutting score on the "moderator" as a basis for eliminating (some) unpredictable cases. Because of space limitation, the reader will need to refer to Hobert and Dunnette for detail. Suffice it to say here that the method leads to a dichotomizing cutting score for optimal classification as to group membership. For the present example, the cutting score becomes 7.5; that is, those in the under-predicted and low-hits groups scoring 8 or 9 are candidates for elimination as being unpredictable.

What of the unpredictables among the high-predictor group? Hobert and Dunnette entertained the idea that items (and scales) that differentiate between the two low-predictor groups may not best differentiate

between the high-hits and overpredicted groups, hence they proceeded to find items (and scales) for a second moderator test. The present author hypothesized that his proposed "moderator," the predictor variable, would do equally well in both situations. It does: $r = .49$ for right-half cases, and the cutting score is 11.5. Thus, those with scores of less than 11.5 (i.e., 11 and 10) become identified as some of the "unpredictables" among the right-half cases.

When one brings together the results of the proposed cutting scores, it is seen that from the entire group those with scores of 8, 9, 10, and 11 on the "moderator" will be eliminated, which is equivalent to sweeping out all the cases in the four middle vertical, shaded, arrays of the scattergram (Figure 1). (Recall that the "moderator" is identical with the predictor.) The percentage eliminated is 45.2, as compared to only 25% in the Hobert and Dunnette study.

But the percentage eliminated is not the ultimate criterion for judging the worth of a moderator. What happens to predictive validity when the moderator is used? This, according to Hobert and Dunnette, may be specified in terms of increased correlation, increased percentage of hits, and decreased percentage of overlap. From the distributions on the *criterion* variable for the two groups, the low predicted and high predicted after elimination by way of the moderator tests, they computed a point-biserial r as .73, whereas the point-biserial for the entire group was .65. For the present "data," it is found that the point-biserial is increased from .55 to .72. As for hits, Hobert and Dunnette found an increase of either 4% or 8% (a complication here—perhaps 6% is a good estimate); here an increase of 11% was obtained. For the present example the Tilton (1937) percentage of overlap was reduced from 50 to 30, as compared to a drop from 38 to 28 found by Hobert and Dunnette.

The evident fact that use of the predictor itself as a "moderator" is better than using items or scales selected from a predictor (battery) may or may not come as a surprise. Obviously, less effort is involved: there is no item analysis or search for scales, with

subsequent need for separate cross-validation of the resultant moderator test. However, cross-validation of the predictor battery may still be required, with the cutting scores on the predictor as "moderator" based on the cross-validation group.

The present author's proposal would appear to be a simple breakthrough in methodology of prediction unless one raises the simple question, Do such moderator variables increase accuracy? From the foregoing, the answer seems to be a strong "yes," but one should take a further look. A count of the frequencies in the two hits quadrants of Figure 1 indicates that for *all* cases the percentage of hits is 74.4 as compared to 85.4 for those 548 remaining after use of the "moderator" to eliminate some of the unpredictables. Now of these 548, the total number of hits that can be claimed is 468, or a mere 46.8% of the original starting group. Instead of a gain of 11% there is a loss of 27.6% in the hit rate! It is estimated that Hobert and Dunnette's gain of about 6% is actually a loss of about 15%.

What is going on here is obvious. One can always pull out cases from the central portion of the distribution on the predictor variable. If one dichotomizes on the basis of the predictor both before and after sweeping out such cases, one can expect in general that biserial r will increase, the percentage of overlap will decrease, and if the criterion is also dichotomized the percentage of hits will increase. Or if one dislikes information-wasting dichotomizing, one might calculate the product-moment coefficients. For the present example, r goes up from .687 to .780. But, sad to relate, the error of estimate remains unchanged! This can be seen in Figure 1 by examining the spread of scores within the vertical arrays about the regression line. Neither does the gain in validity claimed by Hobert and Dunnette represent a reduction in the error of estimate despite the increase in correlation.

Anyone who maintains that the apparent gain in hit rate should not be converted to a loss by basing the percentages on the total N should keep in mind that to have the gain one must decline to make predictions for 25%

(Hobert & Dunnette data) or 45% (present "data") of the cases. If the boss will stand for this, maybe he would allow the applied psychologist to attain a 100% hit rate by the simple expedient of refusing to make predictions for 93.2% of the cases when the correlation between predictor and criterion is .70.

REFERENCES

DUNNETTE, M. D. A modified model for test validation and selection research. *Journal of Applied Psychology*, 1963, 47, 317-323.

DUNNETTE, M. D. *Personnel selection and placement*. Belmont, Calif.: Wadsworth, 1966.

GUETZKOW, H., & FOREHAND, G. A. A research strategy for partial knowledge useful in the selection of executives. In R. Taguiri (Ed.), *Research needs in executive selection*. Boston: Harvard Graduate School of Business Administration, 1961.

HOBERT, R., & DUNNETTE, M. D. Development of moderator variables to enhance the prediction of managerial effectiveness. *Journal of Applied Psychology*, 1967, 51, 50-64.

TILTON, J. W. The measurement of overlapping. *Journal of Educational Psychology*, 1937, 28, 656-662.

(Received March 4, 1968)

Manuscripts Accepted for Publication in the

Journal of Applied Psychology

"Test-Wiseness" on Personality Scales: Lawrence J. Stricker*: Center for Psychological Studies, Educational Testing Service, Princeton, New Jersey 08540.

Achievement in Science and Intellectual Style: A. J. Cropley* and T. W. Field: Division of Social Sciences, University of Saskatchewan, Regina, Saskatchewan.

Cognitive, Noncognitive, and Environmental Correlates of Mechanical Ingenuity: W. A. Ownes*: Social Science Research Center, Graduate Studies Research Center, University of Georgia, Athens, Georgia 30601.

Knowledge of Performance as an Incentive in Repetitive Industrial Work: P. S. Hundal*: Department of Psychology, Panjab University, Chandigarh-14, India.

Comparative Study of Need Satisfaction in Governmental and Business Hierarchies: J. B. Rhinehart,* R. P. Barrell, A. S. DeWolfe, J. E. Griffin, and F. E. Spaner: Psychology Service, Veterans Administration Hospital, Coatesville, Pennsylvania 19320.

Predicting Vocational Success for Neuropsychiatric Patients with the Edwards Personal Preference Schedule: Allen Goss*: Baylor University, College of Medicine, 1200 Moursund Avenue, Houston, Texas 77025.

SVIB Interests and Product Preferences: Robert A. Peterson* and Allan L. Pennington: Graduate School of Business Administration, University of Minnesota 55455.

Validity, Predictive Efficiency, and Practical Significance of Selection Tests: Ervin W. Curtis* and Edward F. Alf: 8533 Alpine, La Mesa, California 92041.

* Asterisk indicates author for whom address is supplied.

MULTIDIMENSIONAL ANALYSIS OF WORKER-ORIENTED AND JOB-ORIENTED VERBS¹

JOHN CHARLES ALLEN²

University of Maryland

The purpose of the present study was to determine if a worker-oriented vs. job-oriented (orientation) continuum was unidimensional. Previous research indicated that verbs from job descriptions could be scaled along the orientation continuum. The multidimensional method of successive intervals was applied to a sample of 20 previously scaled verbs. From the responses of 50 college students instructed to consider the verbs in terms of jobs five orthogonal dimensions were obtained. It was concluded that the orientation continuum existed, but that it was too complex to be considered unidimensional. Research based on actual observed job behavior is needed to establish the generality of the orientation continuum and to determine if it is an adequate construct on which to base indirect validity.

McCormick (1964) has attempted to develop an approach to job analysis that would be broadly based, objective, and quantitative. Such an approach could potentially fulfill the demands implied by the concept of synthetic validity (Balma, 1959; Lawshe, 1952; McCormick, 1959). The general idea may also be relevant to the establishment of task taxonomies (Fitts, 1964; Gagné, 1962; Guion, 1965). Much research is necessary, however, before a strategy such as that of McCormick and his associates (McCormick, Cunningham, & Gordon, 1967; McCormick, Cunningham, & Thornton, 1967; Peters & McCormick, 1966) becomes operational.

One way to make the results of job analyses widely applicable and comparable across jobs and industries would be to discover common elements or common denominators among jobs. McCormick (1959) has referred to the end result of such a process as indirect validity. Such terms as generalized validity and synthetic validity (Balma, 1959; Lawshe, 1952) also incorporate the conceptual strategy

used here of applying validity established in one situation to another.

Central to McCormick's approach is the dimension of worker-oriented as contrasted with job-oriented work elements. Job-oriented elements describe what is accomplished by the worker, refer to technological aspects of the job, and include knowledge of the job. Unlike worker-oriented elements, job-oriented elements tend to be specific to a job or a limited class of jobs.

Worker-oriented elements describe in behavioral terms what the worker does to accomplish his objective. Since worker-oriented elements involve commonly observed human behaviors, they are potentially applicable to an unlimited range of jobs and may be useful as common denominators among jobs. A worker-oriented element, using a verb as an example, is LISTENING which might be equally appropriate to describe an element of behavior displayed by such diverse workers as a sonar operator, a piano tuner, or a music critic (Gordon & McCormick, 1962).

The present research was suggested by the results of a study related to McCormick's approach to job analysis and, more specifically, to the semantic aspects of the worker-oriented versus job-oriented (orientation) continuum (Gordon, 1961; Gordon & McCormick, 1962).

The main purpose of the original research by Gordon and McCormick (1962) was to determine if work-related verbs could be

¹ This article is based on a master's thesis completed at the University of Maryland under the direction of C. J. Bartlett while the author was a National Aeronautics and Space Administration Fellow—Fellowship No. NSG(T)-3, #4. Computer time was made available through the facilities of the Computer Science Center of the University of Maryland.

² Requests for reprints should be sent to the author, Department of Psychology, Morrill Hall, University of Maryland, College Park, Maryland 20742.

differentiated along the orientation continuum in a meaningful way. Thus, the original study was primarily intended to contribute to the justification of the use of the continuum with the more complex stimuli studied in subsequent research on job analysis (McCormick, 1964).

Gordon and McCormick (1962) obtained a sample of 1,000 verbs from job descriptions. Through sorting and rating procedures, judges eliminated 700 verbs as inapplicable to the continuum or strictly job oriented. The 300 remaining verbs were rated on a 6-point scale which varied from worker oriented to job oriented. An additional category was provided so that judges could indicate that a verb was ambiguous or fell outside the continuum. The investigators (Gordon & McCormick, 1962; McCormick, 1964) concluded on the basis of the verb study that: (a) The distinction between job-oriented and worker-oriented verbs can be made by both naive and skilled judges with acceptable reliability especially when judgments of a number of judges are pooled; (b) verbs related to human work activities range over a continuum as far as their connotations of worker oriented versus job oriented are concerned; and (c) while the distinction between worker-oriented and job-oriented verbs is not definite, the difference is sufficient to be useful in describing human work in worker-oriented terms.

The procedure of rating the meaning of verbs in terms of the orientation continuum assumed that the meaning of the verbs was unidimensional. That assumption overlooked

the possibility that the meaning of the verbs with respect to their relevance to jobs could be accounted for by more than one dimension. This possibility was supported by the fact that about 20% of the worker-oriented verbs were rated ambiguous by many judges and also by the large dispersion of the ratings of many of the verbs. Further support for the multidimensional nature of words in general comes from the studies performed by Deese (1965), Reeb (1959), and Root (1962). Research by Osgood and his associates (Osgood, Suci, & Tannenbaum, 1957) and Andrews and Ray (1957) supports the idea that the judgment of words is usually multidimensional.

For these reasons some approach for determining the dimensionality of a sample of the verbs used in the original verb study (Gordon & McCormick, 1962) seemed justified. The multidimensional method of successive intervals appeared to be appropriate for this purpose. Discussion of multidimensional scaling procedures, such as those used in the present research, has been presented by Torgerson (1958) and Brown (1967) and will not be repeated here.

The purpose of the present study was to determine whether or not the worker-oriented versus job-oriented continuum could be considered unidimensional. Accordingly, the research hypothesis stated that a multidimensional scaling analysis of a sample of verbs previously scaled along the orientation continuum would identify three types of dimensions: (a) one worker-oriented versus job-oriented dimension, (b) dimensions logically related, but orthogonal to the orientation dimension, such as a dimension incorporating verbs rated ambiguous in the previous study, and (c) orthogonal dimensions not directly related to the orientation dimension such as the relative generality or frequency of the verbs.

METHOD

Stimuli were selected from a list of 300 verbs reported by Gordon and McCormick (1962). Verbs from the extreme ends of the worker-oriented versus job-oriented continuum were selected. Because of differences in frequency count based on the Thorndike-Lorge "G" count (Thorndike & Lorge, 1944), worker-oriented verbs with the lowest frequency and

TABLE 1
VERBS SELECTED FOR USE IN THE PRESENT STUDY

| Job-oriented verbs | Worker-oriented verbs | Ambiguous verbs ^a |
|--------------------|-----------------------|------------------------------|
| BURNS | CONCENTRATES | SCRUTINIZES |
| CURDLES | READS | INTERPRETS |
| COAGULATES | TOUCHES | DEPICTS |
| CARBONIZES | LOOKS | IMAGINES |
| RECEIVES | FEELS | AWAKENS |
| ADHERES | SEES | |
| AGITATES | | |
| ANCHORS | | |
| IMMERSES | | |

^a At least 25% of the judges placed these five worker-oriented verbs in an "ambiguous" category.

job-oriented verbs with the highest frequency were used. Even so, the worker-oriented verbs tended to have higher frequency counts. Where a choice was possible, verbs with smaller dispersions on the orientation continuum were selected. The 20 verbs used in the present study are listed in Table 1.

The 100 male students who participated in the study were fulfilling a course requirement in an introductory psychology course at the University of Maryland. The subjects (Ss) were randomly divided into two groups. Both Group A ($N = 50$) and Group B ($N = 50$) performed the multidimensional scaling task and then rated the verbs in terms of generality. The 5-point rating scale for generality varied from general to specific with only the end points identified.

For both tasks Ss were given booklets with written instructions and asked to work rapidly, but carefully, within a 1-hr. time limit. Effects due to the order of the stimuli in the scaling task were minimized by following a procedure proposed by Ross (1934, 1939). Further precautions were taken to control the position of the stimuli on a page and the location of pages within the booklets. The scaling task was always performed before the verbs were rated on generality.

The first task for Group A was to rate 190 pairs of words constructed from 20 separate verbs on a 7-point scale as a part of the multidimensional scaling procedure. Ratings were made on the basis of similarity as perceived by the rater. No specific context or set was introduced. The Ss were familiarized with the verbs by requiring them to select pairs of verbs to serve as anchors for the extreme ends of the similarity scale. Following the successive-intervals task, Ss were requested to rate the 20 verbs on the generality scale.

The instructions used with Group A were supplemented in Group B by instructing Ss to consider the stimulus pairs in relation to actions performed in "jobs in general." First, Ss were required to rate 190 pairs of verbs on a 7-point scale on the basis of similarity. The importance of the job context was stressed in the instructions. Then, Ss rated 20 verbs on the 5-point general versus specific scale as in Group A.

RESULTS

The raw data for the multidimensional scaling procedure were obtained by requiring each S to rate 190 stimulus pairs along a 7-point scale of similarity. A computer program incorporating the Messick and Abelson (1956) additive constant solution was employed to analyze the scaling data. A principal-factor solution and a Varimax rotation were used in the solutions for both Groups A and B (Harman, 1960; Kaiser, 1958).

For Group B, given job-context instructions, a break in the eigenvalues was used to

TABLE 2
ROTATED FACTOR MATRIX FOR THE MULTI-DIMENSIONAL SCALING ANALYSIS
IN GROUP A

| Stimulus | Factor | | | | |
|--------------|--------|-------|------|-------|-------|
| | I | II | III | IV | V |
| SCRUTINIZES | .76 | .47 | -.39 | .13 | .10 |
| CARBONIZES | -.16 | .02 | .54 | -1.97 | .00 |
| CONCENTRATES | .88 | -.03 | -.08 | .09 | -.05 |
| RECEIVES | -.01 | -.30 | -.47 | .16 | .03 |
| INTERPRETS | .44 | -.06 | -.66 | .36 | .63 |
| ADHERES | -.34 | -1.33 | .06 | .07 | -.09 |
| READS | 1.18 | .16 | -.23 | .47 | .27 |
| CURDLES | -.20 | -.19 | 1.77 | -.42 | -.21 |
| DEPICTS | .20 | .20 | -.36 | .41 | .68 |
| AGITATES | -.94 | .88 | .86 | -.19 | -.11 |
| IMAGINES | .01 | .38 | -.22 | .96 | .51 |
| COAGULATES | -.02 | -.86 | 1.24 | -.44 | -.01 |
| TOUCHES | -.49 | -.36 | -.35 | -.22 | -.04 |
| BURNS | -1.11 | .55 | .14 | -1.72 | -.05 |
| LOOKS | .55 | .21 | -.50 | .50 | .47 |
| ANCHORS | -.28 | -1.30 | -.11 | -.04 | -1.39 |
| FEELS | -.70 | .17 | -.44 | .24 | .38 |
| AWAKENS | -.47 | 1.02 | -.26 | .76 | -.04 |
| IMMERSES | .28 | .21 | -.01 | -.06 | -1.46 |
| SEES | .44 | .14 | -.52 | .92 | .38 |

Note.—Instructions to S did not specify a job context.

determine the number of dimensions to retain. Then, the same number of dimensions was obtained for Group A. In order to justify this procedure, solutions with varying numbers of dimensions were obtained for Group A. The procedure used to select the number of dimensions to retain in Group A led to only minor variations in the configurations of factor loadings.

Five dimensions were employed to analyze the scaling results for Groups A and B. Based on the eigenvalues and communalities used in the principal-axis solution with no rotations, five dimensions accounted for 48% of the common variance in Group A and 55% in Group B. The rotated factor matrices using five-dimensional solutions for Groups A and B are presented in Tables 2 and 3.

The interpretation of dimensions was limited to results from Group B (see Table 3), since that group was given instructions intended to induce a "job-context" set. Since an orthogonal rotation to a Varimax criterion was used, all of the dimensions were orthogonal. The factors were interpreted and named as follows:

Dimension I—Type of Worker Involvement. This dimension appeared similar to McCormick's orientation continuum. Verbs at the positive pole were from the worker-oriented (see Table 1) and ambiguous categories. The

TABLE 3
ROTATED FACTOR MATRIX FOR THE MULTI-
DIMENSIONAL SCALING ANALYSIS
IN GROUP B

| Stimulus | Factor | | | | |
|--------------|--------|-------|-------|-------|------|
| | I | II | III | IV | V |
| SCRUTINIZES | .57 | .39 | .12 | -.55 | .09 |
| CARBONIZES | -1.96 | -.19 | .57 | .00 | -.04 |
| CONCENTRATES | .80 | -.18 | .63 | -.14 | .16 |
| RECEIVES | .10 | .43 | -.71 | -.08 | -.20 |
| INTERPRETS | .82 | -.20 | -.13 | -.58 | -.50 |
| ADHERES | .02 | -1.34 | -.16 | .70 | .16 |
| READS | .94 | -.07 | .09 | -.53 | -.09 |
| CURDLES | -.67 | .09 | 1.13 | 1.33 | -.18 |
| DEPICTS | .26 | -.10 | -.06 | -1.00 | -.65 |
| AGITATES | -.79 | 1.28 | -.06 | .56 | .11 |
| IMAGINES | .31 | .15 | .14 | -.98 | -.22 |
| COAGULATES | -.46 | -.61 | 1.19 | .90 | -.32 |
| TOUCHES | -.19 | -.04 | -1.01 | .27 | -.03 |
| BURNS | -1.82 | .58 | -.18 | .29 | .24 |
| LOOKS | .83 | .55 | -.06 | -.31 | -.25 |
| ANCHORS | .10 | -1.22 | -.12 | .33 | 1.26 |
| FEELS | .00 | .13 | -.78 | -.08 | -.09 |
| AWAKENS | .57 | 1.04 | -.67 | .18 | -.25 |
| IMMERSES | -.23 | -.08 | .25 | .18 | 1.30 |
| SEES | .79 | .24 | -.18 | -.49 | -.50 |

Note.—Instructions to S specified a job context.

verbs were READS, LOOKS, INTERPRETS, CONCENTRATES, and SEES. Words at the opposite pole were from the job-oriented category. These included CARBONIZES, BURNS, AGITATES, and CURDLES.

Dimension II—Actions Related to Motion. All of the words with high loadings on this dimension except AWAKENS were job-oriented verbs. The dimension contrasted AGITATES and AWAKENS with ADHERES, ANCHORS, and COAGULATES.

Dimension III—Degree of Personal Involvement in Actions. When considered as a bipolar continuum, the dimension related to actions involving people or personal contact as opposed to actions involving impersonal manufacturing processes. Verbs with negative loadings were TOUCHES, FEELS, RECEIVES, and AWAKENS; verbs with positive loadings were COAGULATES, CURDLES, and CONCENTRATES. Note that RECEIVES occurred with the worker-oriented rather than the job-oriented verbs as expected. The same type of reversal occurred with CONCENTRATES, a worker-oriented verb.

Dimension IV—Degree of Mental Activity. The negative pole had high loadings on DEPICTS and IMAGINES, while on the positive pole CURDLES, COAGULATES, and ADHERES were found. In contrast to Dimension I, where both worker-oriented and ambiguous verbs ap-

peared at the same pole, this dimension had only ambiguous verbs at one pole.

Dimension V—Marine Activities. Dimension V had high loadings on IMMERSES and ANCHORS. At the opposite pole, DEPICTS had a moderate loading. Unlike the other dimensions, Dimension V seemed to be unipolar.

The reliability of the multidimensional scaling data was evaluated with a variation of the split-half technique. Groups A and B were randomly divided into two subgroups of 25 Ss each. For both subgroups, the median similarity rating given each verb pair was determined. Then, the reliability estimate was obtained by correlating the median item ratings for the sample halves and making the usual split-half correction for length. This procedure was repeated 20 times for Groups A and B to get mean values.

The introduction of the job-context instructions resulted in increased average reliability of the similarity ratings from .55 in Group A (no job context) to .73 in Group B (job context). On the basis of this increase in reliability, it was inferred that the job-context instructions actually affected the ratings in the intended direction.

Rank correlations were calculated between the multidimensional scaling dimensions and the generality ratings obtained from Groups A and B. Dimensions III and IV in Group A had correlations exceeding .55, and Dimension III in Group B had correlations exceeding .60. It is of particular interest that neither Dimension I nor IV in Group B was highly correlated with the generality ratings. Thus, the dimensions most similar to the orientation continuum could not be explained in terms of the generality or specificity of the verbs.

Ratings of perceived frequency of occurrence of the same 20 verbs were obtained in independent groups similar to Groups A and B. The intercorrelations of frequency with all the dimensions were almost identical with those of the generality ratings. Frequency ratings correlated greater than .85 with the generality ratings, indicating little difference in these two aspects of the sample of verbs.

In contrast to Group B, instructions to Ss in Group A did not mention a job context. The Ss approached the ratings of the verb pairs on similarity with whatever set might

have seemed appropriate to them. The reasons for obtaining data from Group A were to: (a) determine if the job-context instructions in Group B had an effect and (b) use any differences in the factors from Groups A and B to assist in interpreting the results from Group B.

A comparison between the factor loadings given in Tables 2 and 3 for Groups A and B indicated that: (a) In terms of the verbs with the highest loadings, Dimensions II and V were very similar in the two groups; and (b) noticeable differences in a few factor loadings appeared with Dimensions I, III, and IV.

The interrelationships among the dimensions obtained in the solutions for Groups A and B were complex, as indicated by Table 4. The factors that appeared to be most closely related on the basis of both configuration of loadings and rank correlations between factor loadings appear in the same row of that table. Dimension III in Group B had relatively weak relationships with the other dimensions, and no corresponding dimension in Group A is given in Table 4.

DISCUSSION

The first part of the research hypothesis was that a bipolar worker-oriented versus job-oriented dimension would be found. Dimension I (Type of Worker Involvement) in Group B was found to be similar to that continuum and confirmed the hypothesis.

Dimension IV (Degree of Mental Activity) from Group B was relevant to the second part of the research hypothesis. Dimension IV in Group B had high loadings at one pole on verbs referring to mental activities and designated ambiguous in the Gordon and McCormick (1962) study. It had loadings on job-oriented verbs at the opposite pole and had a configuration of stimuli similar to that of Dimension I. The rank correlation between the factor loadings on Dimensions I and IV was $-.70$. The relationship found between these two independent dimensions was interpreted to mean that the orientation continuum was too complex to be represented by a single dimension.

The third part of the hypothesis predicted that the generality or frequency of the verbs would lead to additional dimensions. In Group

TABLE 4
COMPARISON OF DIMENSIONS FROM GROUPS
A AND B

| Corresponding dimension | | Correlation |
|---|---------|-------------|
| Group B | Group A | |
| I Type of Worker Involvement | I | .74 |
| | III | -.71 |
| | IV | .83 |
| II Actions Related to Motion | II | .85 |
| III Degree of Personal Involvement in Actions | | |
| IV Degree of Mental Activity | I | -.66 |
| | III | .72 |
| | IV | -.72 |
| | V | -.83 |
| V Marine Activities | V | -.77 |

Note.—All correlations were Spearman rank correlation coefficients based on $N = 20$.

B, only Dimension III (Degree of Personal Involvement in Actions) was related to these aspects of the verbs. Here the relationship appeared to be moderate as evidenced by a rank correlation in the .60s.

The concepts underlying Dimensions II (Actions Related to Motion) and V (Marine Activities) for Group B seemed to be different from those represented by the other dimensions and were not accounted for by the research hypothesis.

There was a tendency for worker-oriented and job-oriented verbs to fall together at opposite poles of the various dimensions. Exceptions, however, were noted with Dimensions II and III.

The relatively low amount of common factor variance accounted for by five dimensions and the moderate reliability (.73) of the similarity ratings in Group B indicated that the distinction between worker-oriented and job-oriented verbs is not strong. Such a conclusion is in agreement with the earlier research by Gordon and McCormick (1962).

The results of the present study support McCormick's (1959) proposal that a worker-oriented versus job-oriented continuum exists and that verbs can be scaled in terms of the continuum (Gordon & McCormick, 1962). A

new finding was that the perceptions of Ss as measured by the scaling technique involved five dimensions of which two appeared to be generally descriptive of the verb sample. Such an assertion was supported by the finding of a single dimension similar to the orientation continuum, plus the finding of a dimension with high loadings on verbs rated ambiguous in the previous study. Future research may show that the additional dimensions will vary with the context, Ss, and stimulus sample. The demonstration, however, that more than one dimension appeared in the analysis offered sufficient evidence that the worker-oriented versus job-oriented continuum is not unidimensional when verbs are used as stimuli. Thus, while the worker-oriented versus job-oriented dimension exists and it is legitimate to scale such stimuli as verbs along it, the continuum is much more complicated than would be desirable.

A practical problem encountered while selecting worker-oriented verbs for the present study was the lack of a large number of such verbs with small dispersions on the continuum. The scarcity of a large number of worker-oriented elements with which to describe worker activities would seriously impair the usefulness of the worker-oriented concept as a basis for indirect validity in applied situations.

If a consistent theoretical and empirical basis for common denominators applicable to work behaviors or jobs is to be established, the basic units should be widely applicable, unambiguous, relatively free of extraneous influences, and, ideally, unidimensional. The present study raises doubt as to the adequacy of the worker-oriented versus job-oriented elements concept to serve as the common elements upon which to base and infer indirect or synthetic validity. Research using job-analysis data based on actual observed job behaviors in a variety of contexts is needed to establish the generality and value of the orientation continuum. Neither the present research nor the extensive research by McCormick and his associates is sufficient to establish the usefulness of the worker-oriented versus job-oriented continuum.

REFERENCES

- ANDREWS, T. G., & RAY, W. S. Multidimensional psychophysics: A method of perceptual analysis. *Journal of Psychology*, 1957, 44, 133-144.
- BALMA, M. J. The development of processes for indirect or synthetic validity: A symposium. The concept of synthetic validity. *Personnel Psychology*, 1959, 12, 395-396.
- BROWN, K. R. Job analysis by multidimensional scaling. *Journal of Applied Psychology*, 1967, 51, 469-475.
- DEESE, J. E. *The structure of associations in language and thought*. Baltimore: Johns Hopkins Press, 1965.
- FITTS, P. M. Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964.
- GAGNÉ, R. M. (Ed.) *Psychological principles in system development*. New York: Holt, 1962.
- GORDON, G. G. A study of verb meanings in terms of their worker and job orientation. Unpublished master's thesis, Purdue University, 1961.
- GORDON, G. G., & McCORMICK, E. J. A study of the activity connotations of job-related verbs. Report No. 1, September 1962, Purdue University, Occupational Research Center, Contract Nonr-1100(19), Office of Naval Research.
- GUION, R. M. *Personnel testing*. New York: McGraw-Hill, 1965.
- HARMAN, H. H. *Modern factor analysis*. Chicago: University of Chicago Press, 1960.
- KAISER, H. F. The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.
- LAWSHE, C. H. Employee selection. *Personnel Psychology*, 1952, 5, 31-34.
- McCORMICK, E. J. The development of processes for indirect or synthetic validity: A symposium. Application of job analysis to indirect validity. *Personnel Psychology*, 1959, 12, 402-413.
- McCORMICK, E. J. The development, analysis, and experimental application of worker-oriented job variables. Final report, July 1964, Purdue University, Occupational Research Center, Contract Nonr-1100(19), Office of Naval Research.
- McCORMICK, E. J., CUNNINGHAM, J. W., & GORDON, G. G. Job dimensions based on factorial analyses of worker-oriented job variables. *Personnel Psychology*, 1967, 20, 417-430.
- McCORMICK, E. J., CUNNINGHAM, J. W., & THORNTON, G. C. The prediction of job requirements by a structured job analysis procedure. *Personnel Psychology*, 1967, 20, 431-440.
- MESSICK, S. J., & ABELSON, R. P. The additive constant problem in multidimensional scaling. *Psychometrika*, 1956, 21, 1-15.
- OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.
- PETERS, D. L., & McCORMICK, E. J. Comparative reliability of numerically anchored versus job-

- task anchored rating scales. *Journal of Applied Psychology*, 1966, **50**, 92-96.
- REEB, M. How people see jobs: A multidimensional analysis. *Occupational Psychology*, 1959, **33**, 1-17.
- ROOT, R. T. *A multidimensional scaling analysis of meaning*. (Doctoral dissertation, University of Maryland) Ann Arbor, Mich.: University Microfilms, 1962, No. 63-1149.
- ROSS, R. T. Optimum orders for the presentation of pairs in the method of paired comparisons. *Journal of Educational Psychology*, 1934, **25**, 375-382.
- ROSS, R. T. Optimal orders in the method of paired comparisons. *Journal of Experimental Psychology*, 1939, **25**, 414-424.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, Bureau of Publications, 1944.
- TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

(Received February 21, 1968)

(Continued from page 72)

- Influence of Sex Roles on the Manifestation of Leadership: Edwin I. Megargee*: Department of Psychology, Florida State University, Tallahassee, Florida 32306.
- Group and Individual Effects in Problem Solving: George S. Rotter* and Stephen M. Portugal: Psychology Department, Montclair State College, Upper Montclair, New Jersey 07043.
- GRE Psychology Test as an Unobtrusive Measure of Motivation: Robert B. Ewen*: Department of Psychology, New York University, 21 Washington Place, Third Floor, New York, New York 10003.
- Correlates of Job Satisfaction and Job Dissatisfaction among Female Clerical Workers: L. K. Waters* and Carrie Wherry Waters: 45 Eden Place, Athens, Ohio.
- Creativity and Academic Major: Business versus English Majors: Russell Eisenman*: Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.
- Forced-Choice Preferences and Response Latency: Bruce O. Bergum* and Roger P. Dooley: Xerox Corporation, Research Laboratories, 800 Phillips Road, Webster, New York 14580.
- MMPI Scores as Related to Age, Education, and Intelligence among Male Job Applicants: Fred J. Thumin*: School of Business Administration, University of Missouri, 8001 Natural Bridge Road, St. Louis, Missouri 63121.
- SVIB for Women and Demographic Variables in the Prediction of Occupational Tenure: Thomas H. Stone and Gary T. Athelstan*: Assistant Director of Research, Center for Research and Education in the Health Occupations, The American Rehabilitation Foundation, 1800 Chicago Avenue, Minneapolis, Minnesota 55404.

* Asterisk indicates author for whom address is supplied.

FATIGUE AND PERFORMANCE VARIABILITY AMONG TYPISTS

LEONARD J. WEST¹

Division of Teacher Education, City University of New York

Investigation of typewriting fatigue and consistency of performance as a function of level of typing skill employed 234 Ss at typing-skill levels from 5 through 108 wpm. The Ss typed ordinary prose for 30 uninterrupted min. in a manner permitting scoring of the work for speed and number of errors in each minute individually and cumulatively. Although significant fluctuations from a constant level of performance were found at all skill levels, the extent of speed, but not error, fluctuations differed among the skill levels. Principally, decrements in quality of work as the work period progressed decreased with increase in skill level, while correlations between trials and speed for the various skill levels did not exhibit any regular trend. However, absolute decrements in performance during the work period were judged too small to be of any practical consequence and call into question the conventional practice of confining practice and test durations to 1-3 min. during the early months of typewriting training. Relative consistency of speed scores was found to increase with level of skill; however, except for those at the lowest skill levels (who were most inconsistent), relative stability of error scores decreased with increase in skill. Finally, while a 1-min. sample of performance furnishes a highly reliable measure of speed, error measures require at least a 5-min. sample for adequate reliability.

Among the hitherto unanswered questions about work fatigue for a "light, sedentary task" like typewriting are whether (a) increased resistance to fatigue accompanies increases in skill, and (b) increases in skill are accompanied by increased stability of performance across segments of a long, continuous work period. The findings on these questions are intended to contribute to a more complete picture of fatigue phenomena, while findings on the first question have a clear bearing on training practices for typists. These practices appear to be predicated on the supposition that typewriting is a fatiguing task—as inferred from the very short work periods characteristic of early months of training. The tacit assumption seems to be that endurance is a function of high skill and is achieved through a gradual approach to longer durations of continuous work.

These conventional suppositions are somewhat called into question by existing generalizations about fatigue. Ryan (1947), among others, has pointed to negligible fatigue effects in "light, sedentary tasks," with those effects being mostly on quality rather than on quantity of work. Chapanis, Garner, and Morgan (1949) have explained that performance is maintained under fatigue through increased expenditure of energy and increased motivation. In any event, fatigue effects vary with the task.

Concerning typewriting fatigue, Enneis (1956) found no performance differences among employed typists for various typing tasks hypothesized to differ in effortfulness or between manual and electric machine operators. Morgan (1954) found no pulse-rate differences for various typing tasks and various durations of work. Atwood (1964) used slide-camera equipment to photograph the 3- and 10-min. efforts of 30 first-semester typists, finding a drop of 2 wpm after the first minute, but trivial and nonsignificant fluctuations in speed and errors thereafter. Gilmer's (1967) findings for 5- and 10-min. work intervals of 60 first-year typists showed less fluctuation;

¹ This investigation was begun at Southern Illinois University and completed at the City University of New York. The author wishes to thank Barbara Heller for supervising the scoring of the 350,000 words typed by the subjects of this study. Requests for reprints should be sent to the author, Division of Teacher Education, City University of New York, 33 West 42 Street, New York, New York 10036.

an initial decline in speed, followed by an end spurt and a gradual increase in errors throughout each of the two work periods. These patterns were found to hold generally for the various skill levels represented among his Ss.

These earlier investigations employed rather modest durations of continuous work and used Ss at one or another particular level of skill. Accordingly, the primary purpose of the present investigation was to assess typing fatigue over a long (30-min.), continuous work period, as a function of level of skill from novice through expert. Do performance decrements (in speed and quality of work) accompany a long, continuous work period at the typewriter, and do these decrements, if any, vary with skill level? A second purpose of the present investigation was to assess variability or consistency of performance as a function of skill level. Is increasing stability of performance a concomitant of increasing skill? A third purpose of the present investigation, as a by-product of inquiry into fatigue and variability phenomena, was to obtain data on the reliability of measures of typing performance of various lengths. Earlier reliability data were confined to studies using different Ss for tests of various lengths. For ordinary copy work at the typewriter, speed reliabilities under test-retest or parallel-form conditions have typically been in the .80s and .90s for measures as short as 1 min. Error reliabilities, on the other hand, have typically been in the .30s-.40s, even for 5- and 10-min. measures (e.g., Martin, 1954; West, 1956), and only occasionally in the .70s (West & Bolanovich, 1963). The question here is one of optimum length of a single administration: what are the correlations between various cumulative segments of the work period and the full 30-min. work period?

METHOD

Subjects

Of a total *N* of 234 typists, ranging in skill from 5 through 108 wpm, 183 were students in 15 different high school and college typing classes at various stages of training in eight different schools. The remaining 51 persons (at the higher levels of skill) were mainly employed typists, but included a few typewriting teachers and several of the finalists in a national contest for high school typing champions. The relevant population is one of levels of

typing skill (as measured by gross stroking speed during the 30 continuous min. of work of the present investigation), and the sample data are in terms of skill level, not stage of training or amount of work experience.

Test Content and Procedures

The 30 min. of continuous typing involved line-for-line copying of 2,625 words of continuous printed prose, composed so as to be of uniform difficulty at a syllabic intensity (mean number of syllables per word) of 1.40 (representing the conventional estimate of average difficulty of copy materials for vocational typists). Illustratively, the title of this article contains 6 dictionary words and 16 syllables, for a syllabic intensity of 16/6 or 2.67. The materials were printed in triple columns on both sides of an 8½ × 14 in. sheet, long edge horizontal. Accordingly, there was no interruption for turning the copy materials by those who typed less than 43 wpm, one turn for those who typed between 44 and 87 wpm, and only two turns for still faster typists. To preclude interruptions during the work session for changing paper in the machine, paper of sufficient length was cut from teletype rolls.

The Ss were instructed to aim for optimum overall performance, taking both speed and accuracy into account. To permit identification (by the investigator, who administered the testing to all Ss) of each minute of the 30 continuous min. of work without interrupting the typist, the work was done in single spacing, with a double throw of the typewriter carriage upon loud announcement ("throw-throw") by the investigator at the end of each minute. Upon the announcement, Ss were instructed to double space instantly and, without further signal, to continue to type with the *next* line of the printed copy. That tactic substituted a voice stimulus for the usual carriage-throwing stimulus (of ringing bell or visual perception of line end in the copy materials) and was followed, as are the conventional stimuli, by immediate carriage throwing and immediate resumption of typing. The result was 30 sets of single-spaced lines, each set separated from the next by a blank line. To avoid overestimating the work of the first minute and over- or underestimating the work of the last minute of the 30, the first "throw-throw" (signaling the beginning of the first scorable minute) followed 5-15 sec. (half a line) of prior unscored copying. Similarly, the final minute was followed by "throw-throw" and a few seconds of unscored work in a thirty-first minute. In this fashion, individual reaction times to conventional starting and stopping signals were better controlled.

Data and Analysis

Raw data consisted of number of strokes and number of errors per minute, individually and cumulatively. An error was defined as any discrepancy from perfection, but followed the conventional practice of counting no more than one error per word, regardless of number of misstrokes in the

word. (For descriptive purposes here, speed is reported in words per minute, conventionally counting each five typewriter strokes as one word.) The 234 Ss were classified, on the basis of gross stroking speed for the 30 min., into ten 10-wpm skill levels (5-14, 15-24, . . . 85-94, 95-108 wpm) and into four 25-wpm skill levels (5-24, 25-49, 50-74, 75-108 wpm). For purposes of statistical analysis, equal-frequency samples of $N = 160$ (40 from each of the broader ranges) and $N = 171$ (19 from each of nine narrower skill levels, combining 85-108 wpm into one cell) were drawn at random from the total pool of 234 Ss.

On questions of fatigue, performance scores (words per minute and number of errors) were subjected to Levels \times Trials analyses of variance, for 9 (and 4) skill levels for 30 (and 6) trials; that is, 30 individual min. and 6 blocks of successive 5-min. segments of the total 30 min. Interlevel effects were estimated via correlations between trials (30 1-min. and six 5-min. blocks) and mean performance scores for Ss in each of the four 25-wpm skill levels. The conventional assumption that endurance or resistance to fatigue is a function of increasing skill would lead to the expectation of positive, but decreasing, correlations between trials and mean number of errors and negative (and decreasing) correlations between trials and mean speed scores, with increase in skill level.

On questions of consistency of performance, V , the coefficient of variation, was computed for speed and for number of errors for each S , and these V s were subjected to one-way analyses of variance (by skill level). Although the underlying distribution of

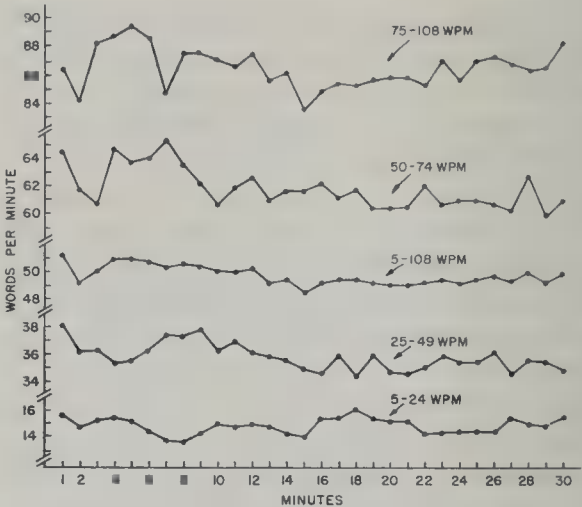


FIG. 1. Mean words per minute in each of 30 consecutive min. of typing, by skill level ($N = 40$ per level).

V is not known, the question at issue demands a measure of relative, not absolute, variability. Use of the standard deviation would necessarily have resulted in larger standard deviations for the faster typists, thereby failing to furnish information on the question of variability in relation to output.

RESULTS AND DISCUSSION

Results are presented, in turn, for (a) fatigue effects, (b) consistency of performance as a function of skill level, and (c) speed and error reliabilities for cumulative portions of the 30-min. work period.

Fatigue Effects

Table 1 contains descriptive data on the performance of all 234 persons, by skill level. The mean of 3.9 (118.15/30) errors per minute (epm) is nearly double the 2 epm found to be characteristic in surveys of the straight copy speeds of students in training (e.g., Robinson, 1967) for 5-min. test durations, and the difference is probably attributable to such factors as Ss' set for the long work period, the novelty of the experimental situation for Ss (instructions to "throw-throw" after each minute), and the knowledge by those Ss who were students that their performance would have no effect on grades.

The minute-by-minute gross words per minute and errors per minute means for $N = 160$ (40 from each of 4 broader skill ranges) are displayed in Figures 1 and 2.

Figures 1 and 2 show a slight general trend toward decrements in speed and increments in

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR SPEED AND ERRORS IN 30 MIN. OF CONTINUOUS TYPING, BY SKILL LEVEL

| Skill level (in words per min.) | N | Words per min. | | Total errors | |
|--|-----|-------------------|-------|--------------|-------|
| | | M | SD | M | SD |
| 5-14 | 40 | 10.78 | 2.14 | 64.98 | 27.39 |
| 15-24 | 25 | 18.88 | 2.83 | 99.12 | 36.44 |
| 25-34 | 34 | 28.35 | 2.36 | 150.50 | 72.20 |
| 35-44 | 19 | 38.32 | 2.65 | 149.84 | 75.35 |
| 45-54 | 21 | 49.76 | 2.28 | 115.86 | 34.44 |
| 55-64 | 26 | 59.42 | 3.04 | 128.07 | 74.52 |
| 65-74 | 27 | 69.22 | 3.03 | 132.14 | 83.21 |
| 75-84 | 20 | 78.65 | 3.00 | 109.95 | 50.52 |
| 85-94 | 13 | 88.53 | 3.26 | 101.77 | 45.71 |
| 95-108 | 9 | 99.89 | 4.20 | 102.00 | 43.79 |
| Total | 234 | | | | |
| Grand M s and SD s ^a | 171 | 49.77 | 26.85 | 118.15 | 64.87 |

^a These are based on a randomly selected 19 persons from each of nine skill levels, combining 85-108 wpm into one cell, that is, on distributions of speed and of error scores for $N = 171$.

errors as the work period proceeds. By no means are the changes large in an absolute sense, nor is there any abrupt shift toward decreased speed or increased errors. Mainly, the fluctuations from minute to minute in both speed and errors show that any cumulative fatigue for this typewriting task is, at most, faint. Instead, the fluctuations would appear to reflect temporary changes in energy expenditure as a means of managing and at least partly recovering from whatever fatigue may have accumulated during portions of the work period.

Analyses of variance for speed and for errors were carried out for the four broader (25-wpm) and nine narrower (10-wpm) skill levels across both 30 (1-min.) trials and 6 (successive blocks of 5-min.) trials, with results as shown in Table 2.

As shown in Table 2, the eight obtained F s for trials were uniformly significant ($p < .01$), showing, for both broad and narrow skill ranges for both 1-min. and 5-min. segments of the total 30-min. work period, that speed and error scores departed significantly from an identical level throughout the work period. The obtained F s for the Levels \times Trials interaction (significant for speed but not for errors) show that speed fluctuations varied from one skill level to another, whereas error fluctuations were similar across all levels of typing skill. (Findings on relative variability on a level-by-level basis are discussed later.)

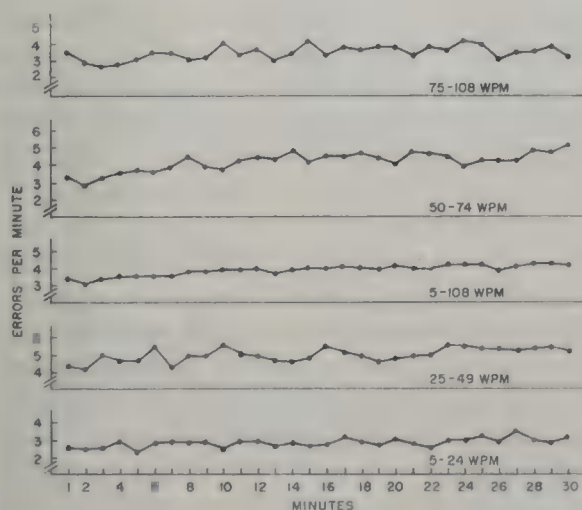


FIG. 2. Mean errors per minute in each of 30 consecutive min. of typing by skill level ($N = 40$ per level).

TABLE 2

ANALYSIS OF VARIANCE FOR SPEED AND ERRORS FOR FOUR COMBINATIONS OF SKILL LEVELS AND TRIALS

| Source | df | I. 4 25-wpm levels and 6 5-min. trials | | | |
|--------------|-----|--|---------|--------|--------|
| | | Speed | | Errors | |
| | | MS | F | MS | F |
| Levels | 3 | 5833646.60 | 712.41* | 201.52 | 7.88* |
| Error | 156 | 8188.55 | | 25.54 | |
| Trials | 5 | 1214.00 | 11.50 | 14.62 | 12.30* |
| L \times T | 15 | 435.40 | 4.12* | 1.56 | 1.31 |
| Error | 780 | 105.49 | | 1.18 | |
| Total | 959 | | | | |

II. 4 25-wpm levels and 30 1-min. trials

| | | | | | |
|--------------|------|-------------|---------|---------|-------|
| Levels | 3 | 29168240.00 | 712.41* | 1007.63 | 7.88* |
| Error | 156 | 40942.50 | | 127.71 | |
| Trials | 29 | 1844.13 | 4.72* | 14.79 | 3.55* |
| L \times T | 87 | 1104.71 | 2.82* | 4.62 | 1.11 |
| Error | 4524 | 390.48 | | 4.15 | |
| Total | 4799 | | | | |

III. 9 10-wpm levels and 6 5-min. trials

| | | | | | |
|--------------|------|------------|----------|-------|--------|
| Levels | 8 | 2260608.70 | 1332.80* | 82.72 | 3.26* |
| Error | 162 | 1696.12 | | 25.37 | |
| Trials | 5 | 1521.00 | 15.02* | 16.01 | 13.21* |
| L \times T | 40 | 300.25 | 2.96* | 1.33 | 1.09 |
| Error | 810 | 101.20 | | 1.21 | |
| Total | 1025 | | | | |

IV. 9 10-wpm levels and 30 1-min. trials

| | | | | | |
|--------------|------|-------------|----------|--------|-------|
| Levels | 8 | 11303045.00 | 1332.78* | 413.60 | 3.26* |
| Error | 162 | 8480.74 | | 126.86 | |
| Trials | 29 | 2140.68 | 5.67* | 16.29 | 3.90* |
| L \times T | 232 | 884.31 | 2.26* | 4.02 | .96 |
| Error | 4698 | 390.71 | | 4.17 | |
| Total | 5129 | | | | |

Note.— N s of 160 for four levels and of 171 for nine levels.

* $p < .01$.

Mean performance scores for the analyses of Section I of Table 2 are displayed in Tables 3 and 4. The trend of these scores across trials may be taken as representative of those underlying the analyses of Sections II, III, and IV of Table 2.

TABLE 3

GROSS WORDS PER MINUTE IN SIX SUCCESSIVE 5-MIN. SEGMENTS OF A 30-MIN. WORK PERIOD, BY SKILL LEVEL

| Skill level | Min. | | | | | | |
|-------------|------|------|-------|-------|-------|-------|------|
| | 1-5 | 6-10 | 11-15 | 16-20 | 21-25 | 26-30 | 1-30 |
| 5-24 | 15.2 | 14.1 | 14.5 | 15.4 | 14.4 | 14.9 | 14.8 |
| 25-49 | 36.2 | 37.0 | 35.8 | 35.0 | 35.2 | 35.2 | 35.8 |
| 50-74 | 62.9 | 63.3 | 61.7 | 61.1 | 61.0 | 60.8 | 61.8 |
| 75-108 | 87.3 | 87.1 | 85.9 | 85.4 | 86.2 | 87.1 | 86.5 |
| 5-108 | 50.4 | 50.4 | 49.5 | 49.4 | 49.2 | 49.5 | 49.7 |

Note.— $N = 40$ per level.

TABLE 4
ERRORS PER MINUTE IN SIX SUCCESSIVE 5-MIN.
SEGMENTS OF A 30-MIN. WORK PERIOD,
BY SKILL LEVEL

| Skill level | Min. | | | | | | |
|-------------|------|------|-------|-------|-------|-------|------|
| | 1-5 | 6-10 | 11-15 | 16-20 | 21-25 | 26-30 | 1-30 |
| 5-24 | 2.55 | 2.74 | 2.75 | 2.86 | 2.92 | 3.08 | 2.82 |
| 25-49 | 4.51 | 4.97 | 4.77 | 4.96 | 5.23 | 5.28 | 4.95 |
| 50-74 | 3.34 | 3.92 | 4.44 | 4.63 | 4.46 | 4.66 | 4.24 |
| 75-108 | 2.99 | 3.47 | 3.62 | 3.77 | 3.90 | 3.56 | 3.55 |
| 5-108 | 3.35 | 3.77 | 3.89 | 4.05 | 4.13 | 4.14 | 3.89 |

Note.—*N* = 40 per level.

Despite the significant *F*s for trials and for interaction (Table 2), the among-trials differences in speed shown in Table 3 are trivially small, both within and across levels. Intertrial differences within levels are as small as .0 and .1 wpm; across levels (5-108 wpm) no change in speed from one 5-min. segment to any other exceeds 1.2 wpm. The statistically significant changes in speed among trials are felt to be of little practical consequence. The data of Table 4, on the other hand, reveal progressive increments in errors as the work period proceeds, both within and across skill levels, thus supporting the existing generalization about fatigue effects on quality of work in light, sedentary tasks. At the same time, intertrial differences as small as .01 and .06 epm were found (within levels), while across levels (5-108 wpm) no change in accuracy from one 5-min. segment to any other exceeds .8 epm. In general, the absolute size of the speed and error fluctuations among trials is judged not to justify the conven-

TABLE 5
CORRELATIONS BETWEEN TRIALS AND
PERFORMANCE MEANS

| Skill level | Speed | | Errors | |
|-------------|-----------|----------|-----------|----------|
| | 30 trials | 6 trials | 30 trials | 6 trials |
| 5-24 | .033 | .033 | .648 | .976 |
| 25-49 | -.595 | -.782 | .603 | .896 |
| 50-74 | -.656 | -.909 | .760 | .872 |
| 75-108 | -.134 | .289 | .538 | .728 |
| 5-108 | -.616 | -.830 | .869 | .923 |

Note.—*N* = 40 per level.

tional restriction of practice durations during training to just a few minutes—not for reasons of supposed substantial fatigue.

Interlevel differences. On the issue of whether increased resistance to fatigue accompanies increases in skill, the obtained correlations between trials (30 1-min. and 6 5-min. blocks) and mean performance scores of the 40 Ss within each of four 25-wpm skill ranges are shown in Table 5.

The magnitudes of the correlations across all Ss (5-108 wpm), shown in the bottom row of Table 5, make apparent the trend toward performance decrements as trials accumulate—more markedly for errors than for speed. On the issue of differential effects for speed, the correlations for the least and the most skilled typists do not differ significantly from zero. More pertinent (for six trials), while these two correlations differ significantly from those for the two middle levels of skill ($p < .001$), there is no apparent trend in the speed correlations with increase in skill level. The absence of progressive differential effects on speed of work accompanying increases in skill is in accord with the existing generalization about negligible fatigue effects on speed of work in light, sedentary tasks.

For errors, on the other hand, the hypothesized progressive decrease in correlations with increase in skill has some, if not consistent, support in the data (i.e., for 5-min., but not for 1-min., trials). Specifically, for six trials, interlevel differences in the obtained error correlations are statistically significant (p 's ranging from $< .05$ through $< .001$) between all levels except between the two middle and between the two highest levels. It seems likely that skill levels identified as 5-24, 25-49, and 50-108 wpm would be found to differ from each other with respect to correlations between skill level and mean errors per minute scores. In general, then, if not in clear step-wise fashion for the four skill levels of the present investigation, as skill increases there does appear to be increasing resistance to decrements in quality of work during a long work period.

Consistency of Performance

For the total sample of 234 Ss, unequally distributed among four 25-wpm skill levels

(numbered 1–4 from low to high), mean speed V s of 15.20, 9.95, 7.39, and 5.72 were found, respectively, and the obtained F in one-way analysis of variance for these V s ($df = 3/230$) is highly significant ($p < .001$). Consistency of performance varies regularly with skill level. Although the plots of Figure 1 suggest greater absolute speed variability among more highly skilled typists, the decrease in V s as skill level increases shows that *relative* consistency in speed of performance increases with skill—a finding that is in accord with intuitive expectations. The reverse was found, with one important shift in rank order, for errors. Although the obtained F was highly significant ($p < .01$ for $df = 3/230$), showing that relative variability in errors does vary with skill level, the V s for the four skill levels (in 1–4 order) were 64.8, 47.6, 56.6, and 60.8. The skill levels are in 1-4-3-2 order from least to most consistent in relative accuracy. The greatest inconsistency among least skilled typists might be expected on the grounds of the notorious variability in work methods (stroking techniques) exhibited by novices at perceptual motor skills involving fine movements and control over the small muscles. The apparent trend thereafter toward decreasing consistency, increasing variability, in relative accuracy with increase in skill is puzzling and contrary to intuitive expectations. In summary, on the question of consistency of performance with increase in skill, in accordance with expectations speed grows more stable with increase in skill, whereas no regular trend was found for relative accuracy; in fact, for those above novice levels of skill, the trend toward decreasing consistency in relative accuracy is contrary to normal expectations.

Score Reliability

Intercorrelations among speed and among error scores were computed for cumulative portions of the 30-min. work period (1, 2, 3, 5, 10, 15, 20, and 30 min.). The intercorrelation matrix for speed (not shown here) reveals no r below .985. For a single measure, it is clear that a 1-min. measure furnishes, for all practical purposes, as reliable an index of speed as is provided by substantially longer tests. The reliability of error measures in

TABLE 6

ERROR INTERCORRELATIONS AMONG CUMULATIVE
SEGMENTS OF A 30-MIN. WORK PERIOD
AT THE TYPEWRITER

| Cumulative min. | Cumulative min. | | | | | | |
|-----------------|-----------------|------|------|------|------|------|------|
| | 2 | 3 | 5 | 10 | 15 | 20 | 30 |
| 1 | .858 | .808 | .753 | .695 | .669 | .652 | .652 |
| 2 | | .951 | .900 | .824 | .794 | .769 | .747 |
| 3 | | | .951 | .883 | .855 | .826 | .801 |
| 5 | | | | .954 | .926 | .903 | .877 |
| 10 | | | | | .980 | .961 | .933 |
| 15 | | | | | | .989 | .967 |
| 20 | | | | | | | .985 |

Note.— $N = 234$.

short work samples is decidedly lower, as shown in Table 6.

If reliabilities at least in the .90s should be sought provided they are achievable in tests of practicable length, it appears from the data of Table 6 that a reasonably reliable estimate of typewriting errors cannot be achieved in less than 5 min. The 1-min. test timings characteristic of the early weeks of training and the 3-min. test timings characteristic of the next few months provide measures of errors whose reliability is too low to make those measures usable. For any single measure, it would appear that a 5-min. duration should be a minimum. This is not to say that a single measure is adequate; for, as mentioned earlier, typewriting errors vary widely from one testing occasion to another. However, if a single measure is used, 5 min. should be a minimum duration.

Implications for Training and Employment Testing

Actually, an accumulation of evidence summarized by West (1967) casts doubt on the propriety of the heavy focus on ordinary copying skills in training and in employment testing for typists. At the same time, until teachers and employers come to appreciate the substantial irrelevance of "straight copy" skills to proficiency at realistic typing tasks, it would appear desirable to conduct that training and testing with maximum efficiency. The findings of the present investigation suggest that there is no reason to confine much

of early training to practice durations of 1-3 min. The stimulus-response conditions that lead to maximum positive transfer mandate as close as possible a match between the practice durations of training and those of employment testing. While a 1-min. sample of straight copy performances furnishes a sufficiently reliable measure of pure stroking speed, an acceptably reliable index of stroking errors in a single test requires at least a 5-min. measure. Accordingly, 5-min. practice and test durations seem advisable for straight copy typing skills.

REFERENCES

- ATWOOD, D. D. Periodicity of effort in beginning typewriting. Unpublished doctoral dissertation, University of North Dakota, 1964.
- CHAPANIS, A., GARNER, W. R., & MORGAN, C. T. *Applied experimental psychology*. New York: Wiley, 1949.
- ENNEIS, W. H., JR. *Effect of machines, tasks, and tiredness feelings on various measures of output curves of typists*. (Doctoral dissertation, Pennsylvania State University) Ann Arbor, Mich.: University Microfilms, 1956, No. 56-2010.
- GILMER, L. L. The effects of fatigue on speed and errors. *Typewriting News*, Fall 1967, 1, 3.
- MARTIN, G. E. *The effects of using continuous and interval speed-forcing programs in learning to typewrite*. (Doctoral dissertation, University of Pittsburgh) Ann Arbor, Mich.: University Microfilms, 1954, No. A54-3322.
- MORGAN, R. W. *The measure of stress and strain of timed writings and production typing as reflected in pulse rates*. (Doctoral dissertation, University of Pittsburgh) Ann Arbor, Mich.: University Microfilms, 1954, No. A54-3411.
- ROBINSON, J. W. The relation of copy difficulty to typewriting performance. *Delta Pi Epsilon Journal*, 1967, 9 (2), 25-30.
- RYAN, T. A. *Work and effort*. New York: Ronald Press, 1947.
- WEST, L. J. An experimental comparison of nonsense, word, and sentence materials in early typing training. *Journal of Educational Psychology*, 1956, 47, 481-489.
- WEST, L. J. Production proficiency among typists—research and implications. *Business Education Forum*, 1967, 22 (2), 5-7.
- WEST, L. J., & BOLANOVICH, D. J. Evaluation of typewriting proficiency training: Preliminary test development. *Journal of Applied Psychology*, 1963, 47, 403-407.

(Received February 28, 1968)

**POSITION FOR PSYCHOLOGIST
WITH
WORK EXPERIENCE IN BUSINESS AND INDUSTRY
TO WORK IN
THE APPLIED-RESEARCH PROGRAM
OF
THE SMITH RICHARDSON FOUNDATION'S
CREATIVE LEADERSHIP LABORATORY**

WE SEEK A PSYCHOLOGIST who holds the Ph.D. whose background includes practical professional experience in the business setting.

THIS APPLIED-RESEARCH POSITION requires a high degree of skill and interest in the diagnosis and evaluation of normal personality patterns as a basis for predicting managerial or leader behavior.

THE POSITION IS PART OF A PROGRAM which seeks to evaluate, improve and teach the process of psychological evaluation as a means of identifying creative leadership talents. It will involve research, teaching and applied practice.

SALARY OPEN, dependent upon background and experience.

APPLICANTS should write to:

**James N. Farr, Ph.D.
Technical Director
Smith Richardson Foundation, Inc.
Box 3265
Greensboro, North Carolina 27402**

NEW IN 1969

DEVELOPMENTAL PSYCHOLOGY

Editor: Boyd R. McCandless

Developmental Psychology will contain articles which represent the broad range of growth and development and their major associated variables. Chronological age as well as sex, socioeconomic status and efforts of physical growth variables (including time of maturation and body build) are all considered relevant development variables.

Although the adolescent and the aging population appear to deserve particular attention at this time, *Developmental Psychology* will indeed cover the entire span. Cross-species articles and articles concerning developmental research with retardates will also be included.

The articles in *Developmental Psychology* will not be limited to any one approach such as "experiments versus studies" or "cognition versus motivation".

**Bimonthly. One volume per year.
Subscription Price: \$10.00 per year**

**AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N. W.
Washington, D. C. 20036**

journal of applied psychology

Table of Contents

| | | |
|--|--|----------|
| Organizational Factors and Individual Performance: A Longitudinal Study. | George F. Farris | 87 |
| Conceptual and Operational Problems in the Measurement of Various Aspects of Job Satisfaction. | Martin G. Evans | 93 |
| Convergent and Discriminant Validities between the Cornell Job Descriptive Index and a Measure of Goal Attainment. | Martin G. Evans | 102 |
| A Study of Some EPPS Variables as Factors of Academic Achievement. | R. P. Bhatnagar | 107 |
| Experienced and Prior Probability in a Complex Decision Task. | Michael H. Strub | 112 |
| Punitive Supervision and Productivity: An Experimental Analog. | David R. Schmitt | 118 |
| Marginal Productivity Procedure for Staff Selection. | Purnell H. Benson | 124 |
| Achievement in Science and Intellectual Style. | A. J. Cropley and T. W. Field | 132 |
| Pattern Analysis of Biographical Predictors of Success as an Insurance Salesman. | Robert Tanofsky, R. Ronald Shepps, and Paul J. O'Neill | 136 |
| Brand Awareness: Differential Roles of Fittingness and Meaningfulness of Brand Names. | Rabindra N. Kanungo | 140 |
| Interaction of Achievement Cues and Facilitating Anxiety in the Achievement of Women. | W. J. McKeachie | 147 |
| Relation Between Birth Order and Being a Beautician. | Philip S. Very and Joseph A. Zannini | 149 |
| Psychological Concomitants and Determinants of Vocational Choice. | Kenneth M. Kunert | 152 |
| Effect of Perceived Expertness upon Creativity of Members of Brainstorming Groups. | Panayiota A. Collaros and Lynn R. Anderson | 159 |
| Validity of Estimates by Clerical Personnel of Job Time Proportions. | Stephen J. Carroll, Jr., and William H. Taylor, Jr. | 164 |
| List of Manuscripts Accepted. | | 106, 151 |

Editor

KENNETH E. CLARE
University of Rochester

Consulting Editors

GENE E. BRUCE
Ohio State University

MARVIN D. DUNNETT
University of Minnesota

NORMAN FREDERIKSEN
Educational Testing Service

LEONARD D. GOODSTEIN
University of Cincinnati

EDWIN R. HENRY
University of Rochester

JOHN HOLLAND
American College Testing
Program

CLIFFORD E. JORGENSEN
Minneapolis Gas Company

DONALD G. MARQUIS
Massachusetts Institute of
Technology

LAURENCE S. MCGAUGHERAN
University of Houston

QUINN MCNEMAR
University of Texas

WARREN T. NORMAN
University of Michigan

HAROLD F. ROTHE
Beloit Corporation

THOMAS A. RYAN
Cornell University

ALEXANDER G. WESMAN
Psychological Corporation

CLARE L. WILSON
Instituto Centroamericano de
Administración de Empresas

This journal gives primary consideration to original investigations in any field of applied psychology except clinical psychology, although a descriptive or theoretical article may be accepted if it represents a special contribution in an applied field. Quantitative investigations of interest or value to psychologists working in the following broad fields will be considered: personnel research in business, industry, and government; engineering psychology; industrial working conditions; research on opinion and morale factors; job analysis and classification research; market and advertising research; vocational and educational prognosis, diagnosis, and guidance at the secondary and college level.

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or they may be obtained from the Editor or from the APA Central Office.

Manuscripts should be addressed to the Editor:

Dr. Kenneth E. Clark, Dean
College of Arts and Science
302 Morey Hall
University of Rochester
Rochester, N. Y. 14627

Authors are strongly advised to follow the general directions given in the *Publication Manual of the American Psychological Association* (1967 Revision). Special attention should be given to the section on the preparation of the references (p. 39). In reference lists, give journal titles in full; do not abbreviate. *All copy must be double spaced, including the references.* All manuscripts should be submitted in *duplicate*, one of which should be an original typed copy; author's name should appear only on title page. Dittoed and mimeographed copies are not acceptable and will not be considered. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail and to check carefully the typing of the final copy.

The following policies govern reprinting or other forms of reproduction of APA-copyrighted journal materials. Permission by the American Psychological Association must be requested for reprinting all tables, figures, and text of more than 500 words (total from one article) in length. There will be a charge of \$5.00 per page for such permission except when the requester is the author of the article to be reprinted, or the article is to be reproduced in a limited number of copies solely for instructional purposes.

Published bimonthly by the
American Psychological Association
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W.
Washington, D. C. 20036

\$10.00 per volume

\$2.00 per issue

HELEN ORR
Managing Editor

CLAIRE PILLIOD
Technical Editor

HENRY S. SALMON
Advertising and
Business Manager

ROBERT D. GORMAN
Subscription Manager

Subscriptions, orders, and business communications should be addressed to the American Psychological Association, 1200 Seventeenth St., N. W., Washington, D. C. 20036. Address changes must reach the subscription office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Second-class postage paid at Lancaster, Pa.

ORGANIZATIONAL FACTORS AND INDIVIDUAL PERFORMANCE:

A LONGITUDINAL STUDY¹

GEORGE F. FARRIS²

University of Michigan

Stability of relationships and time lags in measurement were investigated using information collected at two points in time about organizational factors and the performance of 151 engineers. Four measures of performance were correlated with six organizational factors: involvement in work, influence on work goals, colleague contact, diversity of work activities, salary, and number of subordinates. On the basis of low but statistically significant associations, it was found that correlations between organizational factors and performance were generally stable with a 6-yr. interval between measurements. Surprisingly, relationships were consistently stronger when performance was measured before the organizational factor. It was concluded that changes in organizational factors which follow performance should be considered in research design, organizational theory, and, especially, in interpretations of "simultaneous" associations between organizational factors and performance.

The correlational study done in the field setting is the basis for much of our knowledge about human behavior in organizations. Typically, a questionnaire is administered to individuals to obtain their perceptions of "organizational" factors such as leadership practices or communication, and, at the same time, measurements are made of "output" such as performance or absences. Two assumptions are usually made in such research but rarely tested: (a) the relationships discovered at the particular time of measurement are stable; that is, they would occur for these people in

this organization regardless of the time at which the study was conducted; and (b) the measurements of organizational factors and output refer to essentially the same period of time; that is, the correlations obtained are simultaneous, with no time lag between measurement of the first and second factors.

Because these assumptions are largely untested, they are open to question. Relationships between organizational factors and output may vary over time depending upon such outside circumstances as technology, mission of the organization, external job market, or age of the organization. Findings in the typical single correlational study may be a function of a peculiar combination of such circumstances.

Similarly, the measurements of organizational factors and output may not refer to the same point in time. In most studies people describe the current situation in their organizations (e.g., how satisfied they *are* with their present job), and the output measures are taken over a span of time which is

¹ This paper is based upon the author's dissertation, submitted in partial fulfillment of the requirements for the PhD degree at the University of Michigan. The author is grateful for the comments and suggestions of the members of his committee: Robert L. Kahn, chairman; Frank M. Andrews, Basil S. Georgopoulos, Abraham Kaplan, and J. E. Keith Smith. Part of the research was supported by Grant NSG-489-28-014 from the National Aeronautics and Space Administration.

² Requests for reprints should be sent to the author, Massachusetts Institute of Technology, Sloan School of Management, 52-590, Cambridge, Massachusetts 02139.

apt to include a considerable period prior to the organizational description. This is especially true when the output is infrequent (e.g., patents produced by scientists) or measured subjectively (e.g., top management ratings of divisional performance). In such cases, early as well as current output or impressions are important.

This study investigates the questions of stability and simultaneity using information collected at two different points in time from 151 engineers working for three laboratories of a large electronics corporation. These engineers were among 1,311 scientists and engineers who participated in an extensive investigation of scientists in organizations conducted by Pelz and Andrews (1966). Their investigation, like so many others, was based upon a single measurement of organizational factors and output. No investigation was made of the stability of the relationships obtained, and the output measurements referred to the 5-yr. period preceding the measurement of the organizational factors. One would expect their findings to be especially vulnerable to problems of stability and simultaneity, since the organizations involved were a part of the rapidly changing electronics industry, and the measurements of output referred to a relatively long period of time.

METHOD

Self-report questionnaires were received from the respondents in 1959 and 1965. Included were items asking about six organizational factors and output of patents and reports. In addition, in both 1959 and 1965, colleagues familiar with the respondents' work judged its contribution to science and usefulness to the organization over the past 5 yr. Pearson product-moment correlation coefficients were computed between the factors and performance.

Organizational Factors

Six organizational factors were selected for study because Pelz and Andrews (1966) had found them to be consistently associated with performance. Each was measured on a Likert-type scale. The factors studied were involvement in technical work, influence on work goals, extent of contact with colleagues, diversity of work activities, salary, and number of subordinates.

Performance

Output. Respondents indicated the number of "patents or patent applications" and the number of

"unpublished technical manuscripts, reports, or formal talks (either inside or outside this organization)" which they had produced over the last 5 yr. This information was obtained in both 1959 and 1965. In addition, a question was included in 1965 asking the respondent to report his output for the last 2½ yr. By subtracting responses to this question from those to the previous one, the respondent's output for the first 2½ yr. of the 5-yr. period was determined. Thus, measures of output were available for the time periods 1954–1959, 1960–1965, 1960–1962, and 1963–1965.

Judgments. Senior people from both the supervisory and nonsupervisory levels judged the performance of all respondents with whose work they were directly familiar. They provided rankings of these respondents on two separate measures of performance over the last 5 yr.: contribution to general technical or scientific knowledge in the field and overall usefulness in helping the organization carry out its responsibilities. Because a ratio of one judge for every five respondents was maintained, the work of the great majority of respondents was judged two or more times. Although each judge worked individually, there was substantial agreement among them. These rankings by individual judges were then combined into an overall ranking of all the respondents within a laboratory, and a percentile rank for each respondent on contribution and usefulness was determined following the procedures of Pelz and Andrews (1966, Appendix A).

Adjustment of performance scores. Three factors extraneous to the areas of primary research interest were found when taken together to account for an average of 8% of the variance in the performance scores. They were (a) highest degree earned, (b) time since receiving highest degree, and (c) time with laboratory. Following the procedures of the larger study (Pelz & Andrews, 1966, Appendix C), the performance scores were each adjusted to compensate for deviations from the grand mean of groups at various levels of these three predictor factors.

Predictions. For each relationship between an organizational factor and a measure of performance, it was predicted that a significant positive correlation would occur. These predictions were based on Pelz and Andrews' (1966) earlier findings and supported by several other studies of organizational behavior. Following their earlier procedure, conclusions were drawn according to the pattern of relationships between a given organizational factor and performance. In this study the convention adopted was to display the findings according to the level of statistical significance attained. (Because 60% of the 151 respondents completed a short-form questionnaire in 1959 containing only the questions on involvement and performance, the minimum sample size for the other organizational factors measured in 1959 is 50. Missing data reduced the sample size to 125 for involvement and performance measured in 1959 and all measurements made in 1965.)

TABLE 1

SIZES OF CORRELATIONS NECESSARY FOR SIGNIFICANCE
AT VARIOUS LEVELS OF CONFIDENCE IN THE
PRESENT STUDY

| Level of confidence | Approximate size of r | |
|---------------------|-------------------------|-----------|
| | $N = 50$ | $N = 125$ |
| .01 | .28 | .20 |
| .05 | .22 | .15 |
| .10 | .18 | .13 |
| .15 | .14 | .10 |

Note.— $N = 50$ and $N = 125$ were the two most common sample sizes of this study.

RESULTS

Preliminary Analyses

Before testing the predictions, an examination was made of (a) the interrelationships among the various organizational factors measured at one point in time, (b) the interrelationships among the various performance measures at a single point in time, and (c) the test-retest reliability of the organizational factors and performance from 1959 to 1965. Overall, the organizational factors were not very highly correlated with one another. In 1959 the median correlation was .18, and in 1965, .09. The measurements of performance were also only mildly related with median correlations of .30 and .23 in 1959 and 1965, respectively. The judgments of performance intercorrelated most highly, as one would expect, .63 in 1959 and .56 in 1965, and the measure of reports correlated the lowest with the other measurements. Because of these mild relationships, the four measurements of performance were used separately.

Test-retest reliabilities of the measurements between 1959 and 1965 are shown in Table 2.

The median correlations were .32 for the organizational factors and .46 for performance. The range in reliability of the factors was great, from .10 for contact to .71 for salary. For performance, the range was small, from .39 for patents to .49 for reports. Evidence from another study indicates that the relative instability of these measures reflects changes in the scientist's work situation rather than unreliability in the measuring instruments. Over a 2-mo. interval Pelz and Andrews

TABLE 2

TEST-RETEST RELIABILITIES BETWEEN MEASURES
TAKEN IN 1959 AND 1965

| Measure | r | N |
|----------------------|-----|-----|
| Factors | | |
| Involvement | .46 | 133 |
| Influence | .24 | 51 |
| Contact | .10 | 54 |
| Diversity | .16 | 53 |
| Salary | .71 | 54 |
| No. subordinates | .39 | 56 |
| Performance | | |
| Contribution | .45 | 134 |
| Usefulness | .47 | 137 |
| Patents ^a | .39 | 130 |
| Reports ^b | .49 | 128 |

^a The 1959 measure of patents correlated .39 with patents for the period 1960–1962 and .27 with patents for the period 1963–1965.

^b The 1959 measure of reports correlated .46 with reports for the period 1960–1962 and .43 with reports for the period 1963–1965.

(1966) found a median item test-retest correlation of .62 ($N = 52$) for 89 items from a questionnaire very similar to the one used here. Apparently over the period 1959–1965 there were significant tendencies for those engineers high on performance, salary, involvement, and number of subordinates to continue to be high on these factors. However, previous levels of these factors accounted for only 15–50% of the variance of their levels in 1965. Changes in contact, diversity, and influence were even greater.

Stability of Relationships

The 151 engineers who participated in the present study are a nonrepresentative sample of the 1,311 scientists and engineers of Pelz and Andrews' original investigation. Therefore, the relationships for them between organizational factors and performance in 1959 as well as 1965 were examined. It will be recalled that Pelz and Andrews (1966) found small but consistent positive associations between performance and all the factors under investigation in this study. Table 3 shows the relationship for 151 engineers. It is apparent that in both 1959 and 1965 Pelz and Andrews' (1966) general findings again appear when the engineers of the present study are considered separately. In 1959 significant positive associations occurred between at least one

TABLE 3

RELATIONSHIPS BETWEEN FACTORS AND PERFORMANCE FOR MEASURES TAKEN IN 1959 AND 1965

| Factor | 1959 | | | | 1965 | | | |
|-------------------------|-------------------|-----------------|---------|---------|-------------------|-----------------|---------|---------|
| | Contri- bution | Useful- ness | Patents | Reports | Contri- bution | Useful- ness | Patents | Reports |
| Involvement—self-report | .11* | .18*** | .25**** | — .11 | .11* | .20*** | .29**** | — .18 |
| Influence | .29**** | .16* | .13 | .23*** | .09 | .08 | .19*** | — .03 |
| Contact—number | .02 | .19** | .16* | .05 | — .04 | .07 | — .01 | — .16 |
| Diversity | .15* | — .17 | .23*** | .04 | .05 | — .00 | .11* | .02 |
| Salary | .22*** | .35**** | .22*** | .26*** | .35**** | .42**** | .31**** | .12** |
| No. subordinates | .17** | .47**** | .21*** | .17** | .17**** | .22**** | .16*** | — .00 |

* $p < .15$.
** $p < .10$.
*** $p < .05$.
**** $p < .01$.

measure of performance and all of the organizational factors except contact. (A measure of frequency of contact did show a significant association with patents.) In 1965 significant positive associations occurred between at least one measure of performance and four of the six organizational factors: involvement, influence, salary, and number of subordinates. In the case of diversity, a relationship with patents significant at the .05 level of confidence in 1959 was reduced to a slight tendency (.15 level of confidence) in 1965. In 1959 there were 11 relationships between organizational factors and measures of performance significant at the .05 level. In 1965, there were 9. Thus, although the specific relationships are not identical, two conclusions may be drawn from Table 3: (a) in general, Pelz and Andrews' findings hold for the sample of this study, and (b) the relationships are very similar in 1959 and 1965.

Simultaneity of the Relationships

Relationships between the organizational factors measured in 1959 and performance over the succeeding 5 yr. are shown in Table 4. Four of the organizational factors—involvement, contact, diversity, and the number of subordinates—were related significantly to one kind of subsequent performance. However, these are the only four correlations between the organizational factors and subsequent performance significant at the .05 level of confidence. More than twice as many sig-

nificant relationships occurred in the two situations where performance was measured before the organizational factor (11 in 1959 and 9 in 1965).

A factor-by-factor comparison underscores these differences. For involvement, influence, salary, and number of subordinates, more significant relationships occurred when performance was measured before the factor. For contact the one significant relationship occurred when the factor was measured first, and for diversity the timing of the measurements apparently makes little difference, although performance measured in 1965 related more strongly to the previous level of diversity.

Table 5 shows relationships between the organizational factors and previous and subsequent output for a 2½-yr. period. Again four factors—involvement, influence, diversity, and

TABLE 4

RELATIONSHIPS BETWEEN SIX ORGANIZATIONAL FACTORS AND SUBSEQUENT PERFORMANCE

| Factor | Contri- bution | Useful- ness | Patents | Reports |
|------------------|-------------------|-----------------|---------|---------|
| Involvement | — .33 | .06 | .19*** | — .10 |
| Influence | .18** | .03 | — .00 | .13 |
| Contact | .04 | .22**** | — .05 | — .07 |
| Diversity | .15* | — .15 | .17* | .26*** |
| Salary | .10 | .18** | .01 | .11 |
| No. subordinates | .14* | .25**** | — .08 | .20** |

* $p < .15$.
** $p < .10$.
*** $p < .05$.
**** $p < .01$.

TABLE 5

RELATIONSHIPS BETWEEN ORGANIZATIONAL FACTORS
AND PREVIOUS AND SUBSEQUENT OUTPUT FOR
2½ Yr.

| Factor | Factor measured first | | Output measured first | |
|------------------|-----------------------|---------|-----------------------|---------|
| | Patents | Reports | Patents | Reports |
| Involvement | .18*** | .10 | .23**** | -.15 |
| Influence | .15* | .12 | .17*** | -.04 |
| Contact | .00 | -.03 | -.07 | -.15 |
| Diversity | .29*** | .15* | .21**** | .10* |
| Salary | .13 | .06 | .20*** | .13** |
| No. subordinates | .06 | .21** | .02 | -.01 |

* $p < .15$.

** $p < .10$.

*** $p < .05$.

**** $p < .01$.

salary—show stronger relationships to previous performance. Contact and number of subordinates do not show significant relationships. Apparently, then, the timing of the measurements does make a difference in relationships. When performance is measured first, the relationships are stronger than when the factor is measured first.

In order to determine whether this pattern of associations held under different conditions of measurement, several additional analyses were performed. Separate analyses were conducted for each of the three sites to determine whether peculiarities of the organizational climates of individual laboratories had led to spurious findings. A second approach attempted to minimize the effects of changes in the engineer's job situation so that a person's reported level on a factor in 1959 would be more apt to continue for the entire time span over which subsequent performance was measured. The analysis was repeated for 43 engineers who had been working as "bench scientists" throughout the period of the study (defined as those who had fewer than four subordinates reporting to them in both 1959 and 1965). Third, partial correlations were computed to hold constant past levels of the more recent factor in the zero-order correlations. That is, relationships were determined between each factor and subsequent performance holding constant past performance, and between performance and subsequent amounts

of each factor, holding constant past amounts of each factor. Fourth, the analysis was repeated using eta rather than Pearson r as the measure of association in order to test for curvilinearity. Finally, associations between organizational factors and unadjusted performance scores were examined to determine whether they differed substantially from the associations obtained using the adjusted scores. In each of these additional analyses, the original pattern of findings was strongly supported.³

DISCUSSION

On the basis of low but statistically significant associations, it was found that the relationships between the six organizational factors and performance were stable from 1959 to 1965 but that the assumption of simultaneity did not hold. Relationships were consistently stronger when performance was measured before the organizational factor.

The first finding is encouraging to organizational theorists, since it suggests that relationships between organizational factors and performance are consistent over time, even in so rapidly changing a climate as the electronics industry in the "sputnik era."

The second conclusion, however, is discouraging in two of its implications. It suggests that different conclusions will be drawn in studies relating organizational factors and performance depending on the timing of the measurements. When performance was measured first, stronger relationships occurred than when performance was measured during the period following the measurement of the organizational factor. Most studies of organizational behavior have assumed simultaneity and examined relationships between past levels of performance and the present level of the organizational factor. The assumption of simultaneity, at least for the time spans used in this study, is a doubtful one.

The second implication is that conclusions about causal relationships from such investigations based upon the assumption of simultaneity are apt to be inaccurate. It is logically impossible for a factor which occurred in the past to have been caused by another which

³ For details, see Farris (1967a).

occurs in the present.⁴ Yet organizational theorists looking for patterns of relationships in correlational studies of organizational behavior have drawn such conclusions. For example, on the basis of the extensive body of literature on influence in organizations (e.g., Tannenbaum, 1962), Likert (1961, 1967) has argued that influence causes performance. To the extent that these correlational studies were not based on simultaneous measurements, such a conclusion is unsupported. The more parsimonious causal interpretation is that performance causes influence, since performance was measured over a period of time preceding the measurement of influence.

The most striking finding in this study was that, although four organizational factors were found to relate significantly to one kind of scientific performance, performance was found to relate significantly to at least one subsequent measure of each of the organizational factors. Apparently performance is followed by measurable changes in the social-psychological working environments of people in organizations like those of the present study. Such a phenomenon probably has not

been given sufficient recognition in past theories because it has not often been examined in past research. In the performance-oriented organizations of our society, it has been treated as a desired end result rather than a potential cause. The changes in organizational factors which follow performance should be considered in organizational theory, research design, and interpretations of "simultaneous" associations between organizational factors and performance.

REFERENCES

- FARRIS, G. F. A causal analysis of scientific performance. (Doctoral dissertation, University of Michigan) Ann Arbor, Michigan: University Microfilms, 196, No. 66-14,517, 1967. (a)
- FARRIS, G. F. Toward a nonexperimental method for causal analysis. Working Paper No. 67-279, Sloan School of Management, Massachusetts Institute of Technology, 1967. (b)
- LIKERT, R. *New patterns of management*. New York: McGraw-Hill, 1961.
- LIKERT, R. *The human organization*. New York: McGraw-Hill, 1967.
- PELZ, D. C., & ANDREWS, F. M. *Scientists in organizations*. New York: Wiley, 1966.
- TANNENBAUM, A. S. Control in organizations: Individual adjustment and organizational performance. *Administrative Science Quarterly*, 1962, 7, 236-257.

⁴ Elsewhere the author (Farris, 1967a, 1967b) has developed a method for investigating causal relationships based upon this fact.

(Received February 12, 1968)

CONCEPTUAL AND OPERATIONAL PROBLEMS IN THE MEASUREMENT OF VARIOUS ASPECTS OF JOB SATISFACTION ¹

MARTIN G. EVANS ²

School of Business, University of Toronto

Some of the various concepts and operations that have been suggested for the measurement of job satisfaction are introduced. An effort is made to explore the conceptual and operational relationships between overall job satisfaction, level of aspiration, level of attainment, and level of importance. In particular, note is taken of inappropriate ways in which these three latter concepts have been combined. Finally, a conceptual framework is suggested as a guide to the most appropriate methods of combination.

The concept of job satisfaction is a many-faceted one. Although some students see it as a generalized affective orientation to all aspects of the work situation (Vroom, 1964, p. 99), it is clear that such a view expresses the *resultant* of a whole host of orientations to specific aspects of the job. The respondent, in filling out a measure of general attitude (such as the Brayfield-Rothe, 1951, scale) or in taking an action such as terminating his employment with the organization, is balancing in some complex way the pros and cons of his present job. Students have been quick to realize this and have developed measures that tap various aspects of the job (work itself, supervision, peers, working conditions, and so on). In addition, some have gone further to try to tap the more basic dimensions of a worker's responses about his level of satisfaction of various psychological needs (physiological, safety, social, ego, and self-actualization).

At the outset, some of the different aspects of job satisfaction should be defined in an attempt to provide a consistent vocabulary for use in this paper.

1. The measurement of overall satisfaction. As suggested above, this represents a generalized affective orientation to all aspects of

the job. The methods most frequently used for this purpose involve attitude scales. Such scales may be of the overt, explicit type, such as the Brayfield-Rothe (1951) scale with items like,

I like my job better than the average worker does.

Strongly agree Strongly disagree
or they may be of the projective type such as Kunin's (1955) "Faces" scale. The term "overall job satisfaction" will be used for this concept.

2. The measurement of satisfaction with various aspects of the job. Once again, this is an attitude measurement. Scales may be of the explicit satisfaction type,

The pay I get for my job.

Highly satisfied Highly dissatisfied
or of a descriptive type,

The XYZ Company pays as well as any around here.

Strongly agree Strongly disagree
or of the type such as the Cornell Job Descriptive Index (JDI—Hulin, Smith, Kendall, & Locke, 1963; Kendall, Smith, Hulin, & Locke, 1963; Locke, Smith, Hulin, & Kendall, 1963; Macaulay, Smith, Locke, Kendall, & Hulin, 1963; Smith, 1963; Smith & Kendall, 1963),

Work itself

| | | | |
|-------------|-----|---|----|
| Frustrating | Yes | ? | No |
| Hot | Yes | ? | No |
| Challenging | Yes | ? | No |

The term "job-facet satisfaction" will be used in referring to this concept.

¹ A version of this paper was presented at the Conference of the Association of Canadian Schools of Business in Calgary, June 10, 1968. The author wishes to thank Lee Bolman, Ed Lawler, and Lyman Porter for their stimulating comments.

² Requests for reprints should be sent to the author, School of Business, University of Toronto, 119 St. George Street, Toronto 5, Canada.

3. The measurement of the attainment of either needs or goals. Spitzer (1964, pp. 36–37) tried to make a distinction between the two concepts of need attainment or goal attainment. He saw the latter (goal attainment) as a directly observable behavioral phenomenon and suggested that it can be measured by such items as,

The Opportunity to Develop and Try New Ideas—this means having the chance to be creative by developing new ways to work and supervise. Also to be given the chance to try these new things out so they get a fair test. How much is there now? [Spitzer, 1964, Appendix B].

while the former (need attainment) was seen as a less observable attitude that can only be tapped through the use of a measure that demands a higher level of introspection and self-awareness by the respondent. Spitzer suggested that Porter (1962, 1963a, 1963b) was attempting to make measurements at this level with such items as,

The feeling of self-esteem a person gets from being in my management position.

How much is there now?

However, Spitzer rightly pointed out that many of the scales in Porter's questionnaire are more like goal-attainment items than need-attainment items, for example,

The opportunity for personal growth and development in my management position.

How much is there now?

It would appear that at the operational level the concepts of goal attainment or need attainment have not been distinguished. Accordingly, the term "goal attainment" will be used in describing both concepts.³

4. The measurement of level of aspiration for both needs and goals. This concept deals with the feeling a person has about how much of a particular goal or need *he should have*.

³ It would appear important to make a distinction between this concept (goal attainment) and the preceding one (job-facet satisfaction). Goal attainment (in either of its forms) refers to an estimate by the individual of *how much* of a need or goal he is getting. Job-facet satisfaction refers to *how satisfied* he is with that aspect of the job, in other words, how satisfied he is with his present level of attainment. Note that in operational terms it is possible to distinguish between job-facet satisfaction and goal attainment.

Again, Spitzer's argument about the difference between needs and goals could be applied, but operational separation of the two might prove difficult. The term "goal aspiration" will be used for this concept.

5. The measurement of importance of either (a) job facets or (b) needs and goals. These concepts refer to the saliency of a particular aspect of the job or of a need or goal, that is, the strength of the need or goal to the individual. The terms "job-facet importance" and "goal importance" will be used to refer to these concepts.

Having developed measures for these different aspects of satisfaction, the question arises as to how they should be combined in order to arrive at a job-satisfaction score that best represents the individual's overall affective orientation. To this, there have been several solutions; it is the purpose of this paper to examine them and their implications.

Methods of Combining Aspects of Satisfaction

A review of the literature suggests that at least five ways of combining these aspects to get some measure of overall satisfaction have been developed. They vary in elegance and complexity. It is the aim here to evaluate these combinations. First, however, they must be described.

1. Simple summation of either (a) job-facet satisfaction or (b) goal attainment. The researcher simply asks his respondents about the satisfaction level of each facet or the attainment level of each goal. The total score is obtained by summing over facets or goals.

2. Summation of the product of either (a) job-facet satisfaction and job-facet importance or (b) goal attainment and goal importance. Here the researcher determines how satisfied his respondents are (or how much they have attained) on each dimension and how important each dimension is to them; the level of satisfaction (or attainment) is weighted by the importance, and the product is summed over facets or goals to arrive at an overall score.

3. Summation of the difference between the level of goal aspiration and the level of goal attainment. Here the researcher asks his re-

spondents not only how much of the goal he is now getting, but also how much he thinks that he should be getting; satisfaction is taken to be the difference between aspiration level and attainment level and is summed over goals for the overall score.

4. Summation of the product of goal importance and the difference between level of aspiration and level of attainment. This is similar to Method 3 except that information on importance is also gathered, and the difference score is weighted by this before the final summation takes place.

5. Summation of the differences between goal importance and goal attainment or goal aspiration. Here the researcher gathers information on the two dimensions involved and subtracts from importance the other (aspiration or attainment); the overall score is obtained by summing the differences over goals.

In the discussions that follow, each of the five methods of combination will be discussed in conceptual terms; its appearance in the literature will be reviewed, with special emphasis upon those studies that report the relationship between the combination and a measure of overall job satisfaction. Finally, any possible comparisons with previous combinations will be made.

Simple summation. Conceptually, this is the most simple of the methods of combination. It is, in addition, the easiest of the measures to obtain; only one set of questions need be asked of the respondent. The validity of such a measure depends upon one assumption—that each aspect of the respondent's job-facet satisfaction or goal-attainment space is of equal importance; that is, each can be assigned an equal weight.

Ewen (1967) compared the Cornell JDI, a measure of job-facet satisfaction which taps the facets of work itself, pay, promotion opportunities, supervision, and peers, with the Brayfield-Rothe (1951) scale and the Faces scale (Kunin, 1955), both measures of overall satisfaction. For three samples, the correlations reported in Table 1 are obtained. With both the overall satisfaction scales, the summed JDI shows high correlation (see Table 1). Schaffer (1953) related the sum of need satisfactions (in present terms, a

TABLE 1
CORRELATIONS BETWEEN SUMMED JOB DESCRIPTIVE
INDEX SCORES (UNWEIGHTED) WITH THE
BRAYFIELD-ROTHE AND FACES SCALES

| Sample | B-R | Faces | N |
|--------|-----|-------|-----|
| A | .73 | .74 | 21 |
| B | .50 | .70 | 23 |
| C | .66 | .55 | 120 |

Note.—From Ewen (1967).

measure of goal attainment) to a single measure of overall job satisfaction. The correlation is reported in Table 4. The sum of the attainment of the 12 needs tapped by Schaffer correlated highly with his measure of overall satisfaction. Blai's (1964) study, in which he claimed a correlation of .58 between overall job satisfaction and the sum of need-satisfaction items (measures of goal attainment), is judged to be uninterpretable in that the latter measure showed serious confusion between measures of need satisfaction and measures of need importance; most of the items dealt with importance. As a corollary, Evans (1968) investigated the relationship between job-facet satisfaction and goal attainment. The correlations between common scales are presented in Table 5 for two samples. These are high, and it has been shown that the scales show convergent and discriminant validity (Evans, 1969).

In summary, it would appear that the relationships between overall satisfaction and both job-facet satisfaction and goal attainment are highly positive.

Summation of the product of importance and satisfaction or attainment. Conceptually, this is a more elegant formulation than the first. It takes into account the individual dif-

TABLE 2
CORRELATIONS BETWEEN SUMMED JOB DESCRIPTIVE
INDEX SCORES (WEIGHTED) WITH THE
BRAYFIELD-ROTHE AND FACES SCALES

| Sample | B-R | Faces | N |
|--------|-----|-------|-----|
| A | .75 | .77 | 21 |
| B | .48 | .68 | 23 |
| C | .66 | .56 | 120 |

Note.—From Ewen (1967).

TABLE 3

CORRELATIONS BETWEEN MOST IMPORTANT AND LEAST IMPORTANT JOB DESCRIPTIVE INDEX SCALES WITH THE BRAYFIELD-ROTHE AND FACES SCALES

| Sample | Most important | | | Least important | | |
|--------|----------------|-------|-----|-----------------|-------|-----|
| | B-R | Faces | N | B-R | Faces | N |
| A | .49 | .54 | 21 | .44 | .24 | 21 |
| B | .59 | .63 | 23 | .33 | .41 | 23 |
| C | .61 | .50 | 101 | .34 | .18 | 101 |

Note.—From Ewen (1967).

ferences that may exist in the importance of the facets of satisfaction. In other words, it helps to account for the complex balance that each individual attains in arriving at an overall evaluation of his satisfaction from his reactions to the specific facets of the job. Again, the data are quite simple to gather.

Ewen (1967) weighted the JDI scores (measures of job-facet satisfaction) by the measure of the importance of each scale and correlated the sum of the weighted scales with the Brayfield-Rothe and Faces scales. The results are presented in Table 2. In addition, he compared the correlation of the satisfaction of the most important JDI scale and the least important JDI scale with the overall measures (Brayfield-Rothe and Faces). The results in Table 3 show that the correlations are higher for the most important JDI scale. Schaffer found a similar result for his goal-attainment scales; the results are reported in Table 4. Finally, Ewen obtained very high correlations (.98, .99, .99 for the three samples) between the unweighted total JDI and the weighted (with importance) total JDI.

Comparison among Tables 1, 2, 3, and 4 enable judgments to be made as to whether

TABLE 4

CORRELATION BETWEEN OVERALL JOB SATISFACTION AND NEED SATISFACTION

| Satisfaction | Correlation |
|-----------------|-------------|
| Total | $r = .44^*$ |
| Most important | $r = .54^*$ |
| Least important | $r = .13$ |

Note.—From Schaffer (1953). $N = 72$.
* $p < .001$.

TABLE 5

CORRELATIONS BETWEEN FACETS OF JOB SATISFACTION (JOB DESCRIPTIVE INDEX) AND NEED SATISFACTION

| Sample and facet | r |
|-----------------------|--------|
| Utility ^a | |
| Pay | .59*** |
| Supervision | .60*** |
| Fellows | .25*** |
| Work itself | .29*** |
| Hospital ^b | |
| Pay | .46*** |
| Supervision | .31** |
| Fellows | .23* |
| Work itself | .16* |

Note.—From Evans (1969).
^a $N = 311$.
^b $N = 83$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

the weighting procedure enhances the relationship between overall satisfaction and the sum of the satisfactions with each facet or the goal attainments. The evidence is inconclusive. Both Ewen (job-facet satisfaction) and Schaffer (goal attainment) suggested that slight increases in correlation coefficients can be achieved by weighting the satisfaction components with importance, but such increases are not automatic and in no case are they significant statistically. However, when comparisons are made between least important and most important facets and their correlations with overall satisfaction, the differences are sizable.

Sum of differences in goal-aspiration and goal-attainment levels. This method is slightly more sophisticated than the first. In the first, the researcher is asking the individual how satisfied he is. The individual presumably makes judgments for himself about his aspirations and his present level of attainment with regard to his goals and takes these into account in his answer to the question of how satisfied he is with a particular goal. In this third method, the process is made explicit. The respondent records judgments of his aspiration and attainment; satisfaction is taken to be the difference between them. Thus, in terms of the concepts originally introduced, the following relationship is proposed: For a given job facet and its corresponding goal

areas, job-facet satisfaction is equivalent to the difference between goal aspiration and goal attainment.

Such a combination has had considerable use in the literature. However, it has been used for the operationalization of a number of different concepts. Porter (1962, 1963a, 1963b) referred to it as the “perceived deficiency in need fulfillment,” while Spitzer (1964) used a similar set of operations to measure “goal attainment.” Spitzer reported the relationship between such a scale (in the areas of opportunities to use new ideas, job security, pay, approval from peers, superiors, and subordinates, control over job, promotion, and personal growth opportunities) and the Brayfield-Rothe scale. The multiple correlation coefficient, a statistical device for taking into account all the facets, is presented in Table 6. The correlation is highly positive.

There are no data available to show whether this method of combination is superior to the two previous ones. It has been suggested above that it is conceptually more elegant. With the data collected by Spitzer (1964), a reanalysis could explore the differences between overall satisfaction versus the sum of goal attainments, and overall satisfaction versus the sum of the differences between goal aspiration and goal attainment. A future project is being planned by the author with this in mind.

Sum of the product of importance and the differences. This is similar to the third method except that the difference score is weighted by the importance of the goal before the summation is made. This is conceptually elegant. It makes explicit the differences in importance, in aspiration, and in attainment.

The only data available are those provided

by Spitzer (1964). The highly positive multiple correlation between the Brayfield-Rothe scale and the scale of weighted differences is presented in Table 6.

Comparisons within Table 6 indicate that the fourth method gives a higher multiple correlation than the third method; however, the differences between them are not significant.

Sum of differences between importance and attainment or aspiration. In this method, the sum (over goals) of the differences between responses on importance of a goal and its degree of attainment or level of aspiration is taken to represent the overall satisfaction score. Conceptually this seems meaningless. How can such a difference represent a level of satisfaction? An example will indicate the problems with such a position. Assume three people respond within the following levels of importance and attainment:

| | | | |
|------------|---|---|---|
| Importance | 7 | 3 | 1 |
| Attainment | 7 | 3 | 1 |

They will all, by this method of combination, have equal satisfaction (in this case represented by zero satisfaction). Surely a multiplicative model, $\text{Importance} \times \text{Attainment}$, in which satisfaction scores of 49, 9, 1, respectively, were obtained would be a more accurate representation of reality.

In spite of this, much research has been carried out using this method. In 1960, Glennon, Owens, Smith, and Albright suggested that a method similar to this be adopted for the measurement of morale in order to “permit management to identify ‘sore spots’ or low satisfaction issues [p. 107].” For such a purpose, this measure of combination may be appropriate. It enables management to identify situations in which low satisfaction is coupled with high importance and the opposite situation in which high satisfaction is coupled with low importance. Therefore, it does indicate areas of concern. However, it is not appropriate as a measure of overall job satisfaction.

Empirically, it is conceivable that importance and aspiration level might show a high positive correlation in which case the operations would be interchangeable so that the

TABLE 6

MULTIPLE CORRELATION COEFFICIENT BETWEEN
BRAYFIELD-ROTHE AND FACETS OF NEED
SATISFACTION (WEIGHTED AND
UNWEIGHTED)

| Facet | R |
|------------|------|
| Unweighted | .32* |
| Weighted | .36* |

Note.—From Spitzer (1964).
* $p < .01$.

TABLE 7

MEAN DIFFERENCES IN SCORES (STRENGTH OF NEED,
DEGREE TO WHICH IT IS ATTAINED, DIS-
SATISFACTION) BETWEEN RESIGNED
WORKERS AND MATCHED
CONTINUING WORKERS

| Aspect of need ^a | Type of need | | |
|-----------------------------|------------------|------------------|----------|
| | Recog- nition | Achieve- ment | Autonomy |
| Strength (A) | .08 | .15 | .09 |
| Attainment (B) | -.56*** | -.08 | -.44** |
| Dissatisfaction (A-B) | .64*** | .23* | .53** |

Note.—From Ross and Zander (1957, p. 335). *N* = 169.
^a In all aspects, the scores represent *differences* between
continuing and resigned Workers.
* *p* < .05.
** *p* < .025.
*** *p* < .0025.

combination, importance minus satisfaction, would be a predictor of overall job satisfaction. However, no data have been published that bear on this question. Until such time as it is demonstrated that importance and aspiration are operationally interchangeable, this combination should be avoided as a measure of overall job satisfaction.

Several researchers (Beer, 1966; Kuhlen, 1963; Pelz & Andrews, 1966; Ross & Zander, 1957) have used this combination in order to obtain a measure of overall job satisfaction. Ross and Zander (1957) suggested that leaving the organization (a behavioral manifestation of job dissatisfaction) was associated with dissatisfaction with the fulfillment of the following needs: recognition, achievement,

and autonomy; that is, leavers were more dissatisfied than a demographically matched group that remained with the company. They measured “need strength” (or the importance of the need to the individual) and need satisfaction (the degree to which the job provided fulfillment of the need). The dissatisfaction score was obtained by subtracting need satisfaction from need importance. The results of this study are reported in Table 7. There was little difference between leavers and continuers in need strength (i.e., need importance). In other words, for each need, both leaving and continuing workers rated it equally important. Major differences between the groups were found for the needs of recognition and autonomy in the degree to which the need was satisfied. The resulting intergroup differences in the importance-attainment difference are the direct result (with the possible exception of the difference for the achievement need) of differences in attainment alone. Kuhlen (1963) suggested a strong relationship (for men but not for women) between overall job satisfaction and the discrepancy between potential for need satisfaction in the job and the individual’s need strength (measured on the EPPS, Edwards, 1954). Beer (1966), in order to determine a score for “Actual Need Satisfaction,” obtained the difference between: (a) the score on “Job Inventory” which is a measure of the perceived opportunity to satisfy a need on the job and is equivalent to asking about the extent to which the need is presently satisfied, and (b) the score on “Preference Inventory” which is a measure

TABLE 8
PEARSON PRODUCT-MOMENT CORRELATIONS BETWEEN LEADERSHIP BEHAVIOR AND
WORKER-PERCEIVED OPPORTUNITY FOR NEED SATISFACTION

| Leadership Behavior | Perceived opportunity for need satisfaction | | | | |
|----------------------|---|--------|--------|----------|--------------------|
| | Security | Social | Esteem | Autonomy | Self-actualization |
| Initiating structure | .20* | -.14 | .04 | -.13 | .04 |
| Freedom of action | -.21* | -.16 | .05 | .39** | -.02 |
| Consideration | -.22* | -.20* | .17* | .04 | .21* |
| Production emphasis | .09 | .10 | .01 | -.15 | .06 |

Note.—From Beer (1966, Table 12, p. 43).
* *p* < .05.
** *p* < .01.

TABLE 9

PEARSON PRODUCT-MOMENT CORRELATIONS BETWEEN LEADERSHIP BEHAVIOR AND
WORKER NEED SATISFACTION

| Leadership Behavior | Actual need satisfaction | | | | |
|----------------------|--------------------------|--------|--------|----------|--------------------|
| | Security | Social | Esteem | Autonomy | Self-actualization |
| Initiating structure | .06 | -.08 | .08 | -.02 | -.05 |
| Freedom of action | -.08 | -.19* | .02 | .27* | -.04 |
| Consideration | -.07 | -.21* | .09 | .12 | .12 |
| Production emphasis | .01 | .04 | -.05 | -.07 | -.07 |

Note.—From Beer (1966, Table 13, p. 44).

* $p < .05$.

of need importance. Beer (1966) reported the relationships between leadership behavior (Initiation of Structure and Consideration) and both "Perceived Opportunity for Need Satisfaction" (the extent to which the need is satisfied at present) and "Actual Need Satisfaction" (the hybrid measure under criticism). These results are presented in Tables 8 and 9, respectively. For both measures, the relationships are not strong; however, they do appear to be stronger in the first case where relationships with the conceptually pure measure are presented. Finally, Pelz and Andrews (1966) used as a measure of overall satisfaction the difference between (a) desire, measured by questions about the importance of a particular aspect, and (b) provision, measured by questions about the attainment of a particular aspect. They explicitly (Pelz & Andrews, 1966, pp. 120–121) equated the measure of importance with level of aspiration, a position that, as has been suggested above, requires empirical justification. The results related job satisfaction (and desire and provision) to components of job performance for scientists and engineers. Such a relationship is undoubtedly a complex one, and simple correlations are not to be expected (see Brayfield & Crockett, 1955; Evans, 1969; Spitzer, 1964). However, it would appear that the measure of total satisfaction did not correlate much better with performance than did the total provision score. The correlation of total desire with performance was not strong. It must be pointed out that Pelz and Andrews were aware of some of the difficulties with this method of combination

and preferred that the results for desire and provision be reported separately.

DISCUSSION

In the preceding section, data have been presented that indicate the relationships between overall job satisfaction and the aspects of job-facet satisfaction, goal aspiration, goal attainment, and goal importance. From this data, it is clear that the following combinations have some merit in that they *all* show significant correlations with overall satisfaction:

Combination 1. Overall job satisfaction (JS) is the sum of job-facet satisfaction (JFS) or goal attainment (GAtt).

$$JS = \sum^{\text{facets}} (JFS) \quad [a]$$

$$JS = \sum^{\text{goals}} (GAtt) \quad [b]$$

Combination 2. Overall job satisfaction is the sum of either the product of job-facet satisfaction and job-facet importance (JFI) or the product of goal attainment and goal importance (GImp).

$$JS = \sum^{\text{facets}} (JFS \times JFI) \quad [a]$$

$$JS = \sum^{\text{goals}} (GAtt \times GImp) \quad [b]$$

Combination 3. Overall job satisfaction is the sum of the differences between goal aspiration (GAsp) and goal attainment.

$$JS = \sum^{\text{goals}} GAsp - GAtt$$

Combination 4. Overall job satisfaction is the sum of the product of goal importance and the difference between goal aspiration and goal attainment.

$$JS = \sum_{goals} (GImp \times (GAsp - GAtt))$$

Thus the researcher is faced with a plethora of methods of combination, all of which provide relatively strong correlations with measures of overall job satisfaction. While it may be that decisions about which method to use can be based upon practical expediency (i.e., Ss can only be expected to complete a short questionnaire), it is desirable, given a value system that includes parsimony and elegance in research design, that the decision be made to use a method of combination that is congruent with the researcher's conceptual framework.

With this in mind, the following framework is presented. This, no doubt, has been implicit in much of the earlier discussion, but it should be made explicit at this point. It is an attempt to trace through the logical relationships that exist between the variables.

1. Overall job satisfaction (JS) is a function of the sum (over facets) of the product of job-facet satisfaction (JFS) and job-facet importance (JFI).

$$JS = \sum_{facets} (JFS \times JFI)$$

2. For each facet and its corresponding goals, job-facet satisfaction is a function of the difference between goal aspiration (GAsp) and goal attainment (GAtt).

$$JFS = GAsp - GAtt$$

3. Consequently, overall job satisfaction is a function of the sum (over goals) of the product of goal importance (GImp) and the difference between goal aspiration and goal attainment.

$$JS = \sum_{goals} (GImp \times (GAsp - GAtt))$$

If this conceptual framework is an accurate one, then the researcher can use it as a guide in making his decisions about which method of combination to use. Combinations 2a and

4 are most congruent with the conceptual framework. Combination 4 is probably the most elegant, while Combination 2a combines elegance with brevity of measuring instrument.

One problem remains: the elegant combinations do not appear to be better predictors of overall satisfaction than the others. Why should this be so? One suggestion that can be made here is that the measurement of importance may not be well developed. One observed tendency is for every respondent to report that every goal or facet of the job is of equal importance to him (see Ross & Zander, 1957), with a consequent restriction of the range of variation in the measure. It is suggested that new methods of measuring importance be established.

Finally, it is suggested that Combination 5 (summation of the differences between goal importance and goal attainment or aspiration), which has little conceptual meaning, should be avoided in situations where overall job-satisfaction scores are being computed and where individuals are being compared. As was pointed out above, such a method has its uses; the measurement of overall satisfaction does not appear to be one of them.

REFERENCES

- BEER, M. *Leadership, employee needs, and motivation*. Columbus: Ohio State University, Bureau of Business Research, 1966.
- BLAI, B. An occupational study of job satisfaction and need satisfaction. *Journal of Experimental Education*, 1964, 32, 383-388.
- BRAYFIELD, A. H., & CROCKETT, W. H. Employee attitudes and employee performance. *Psychological Bulletin*, 1955, 52, 396-424.
- BRAYFIELD, A. H., & ROTHE, H. F. An index of job satisfaction. *Journal of Applied Psychology*, 1951, 35, 307-311.
- EDWARDS, A. L. *Edwards Personal Preference Schedule (manual)*. New York: Psychological Corporation, 1954.
- EVANS, M. G. The effects of supervisory behavior upon worker perception of their path-goal relationships. Unpublished doctoral dissertation, Yale University, 1968.
- EVANS, M. G. Convergent and discriminant validities between the Cornell "Job-Descriptive Index" and a measure of goal attainment. *Journal of Applied Psychology*, 1969, 53, 102-106.
- EWEN, R. B. Weighting components of job satisfaction. *Journal of Applied Psychology*, 1967, 51, 68-73.

- GLENNON, J. R., OWENS, W. A., SMITH, W. J., & ALBRIGHT, L. E. A new dimension in measuring morale. *Harvard Business Review*, 1960, 38, 106-107.
- HULIN, C. L., SMITH, P. C., KENDALL, L. M., & LOCKE, E. A. *Cornell studies in job satisfaction: II. Model and method of measuring job satisfaction*. Ithaca: Cornell University, 1963. (Mimeo)
- KENDALL, L. M., SMITH, P. C., HULIN, C. L., & LOCKE, E. A. *Cornell studies in job satisfaction: IV. The relative validity of the Job Descriptive Index and other methods of measurement of job satisfaction*. Ithaca: Cornell University, 1963. (Mimeo)
- KUHLIN, R. G. Needs, perceived need satisfaction opportunities, and satisfaction with occupation. *Journal of Applied Psychology*, 1963, 47, 56-64.
- KUNIN, T. The construction of a new type of attitude measure. *Personnel Psychology*, 1955, 8, 65-77.
- LOCKE, E. A., SMITH, P. C., HULIN, C. L., & KENDALL, L. M. *Cornell studies in job satisfaction: V. Scale characteristics of the Job Descriptive Index*. Ithaca: Cornell University, 1963. (Mimeo)
- MACAULAY, D. A., SMITH, P. C., LOCKE, E. A., KENDALL, L. M., & HULIN, C. L. *Cornell studies in job satisfaction: III. Convergent and discriminant validity for measures of job satisfaction by rating scales*. Ithaca: Cornell University, 1963. (Mimeo)
- PELZ, D. C., & ANDREWS, F. M. *Scientists in organizations: Productive climates for research and development*. New York: Wiley, 1966.
- PORTER, L. W. Job attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology*, 1962, 46, 375-384.
- PORTER, L. W. Job attitudes in management: III. Perceived deficiencies in need fulfillment as a function of line versus staff type of job. *Journal of Applied Psychology*, 1963, 47, 267-275. (a)
- PORTER, L. W. Job attitudes in management: IV. Perceived deficiencies in need fulfillment as a function of size of company. *Journal of Applied Psychology*, 1963, 47, 386-397. (b)
- ROSS, I. C., & ZANDER, A. F. Need satisfaction and employee turnover. *Personnel Psychology*, 1957, 10, 327-338.
- SCHAFER, R. H. Job satisfaction as related to need satisfaction in work. *Psychological Monographs*, 1953, 67(14, Whole No. 364).
- SMITH, P. C. *Cornell studies in job satisfaction: I. Strategy for the development of a general theory of job satisfaction*. Ithaca: Cornell University, 1963. (Mimeo)
- SMITH, P. C., & KENDALL, L. M. *Cornell studies in job satisfaction: VI. Implications for the future*. Ithaca: Cornell University, 1963. (Mimeo)
- SPITZER, M. E. *Goal-attainment, job satisfaction and behavior*. (Doctoral dissertation, New York University) Ann Arbor, Mich.: University Microfilms, 1964, No. 64-10, 10,048.
- VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

(Received February 27, 1968)

CONVERGENT AND DISCRIMINANT VALIDITIES BETWEEN THE CORNELL JOB DESCRIPTIVE INDEX AND A MEASURE OF GOAL ATTAINMENT¹

MARTIN G. EVANS²

School of Business, University of Toronto

The convergent and discriminant validities of the Cornell Job Descriptive Index and the goal-attainment component of Porter's need-satisfaction measure were evaluated according to the Campbell and Fiske (1959) criteria. The coefficient of concordance (*W*) was suggested as a statistical test of the fourth criterion. The scales demonstrated convergent and discriminant validity.

With the plethora of measuring instruments that are available for the measurement of job satisfaction, it is important that the relationships between such measures be reported so that the investigator, wishing to use one of these measures, can be aware of its current status. This article presents the relationship between a well-documented measure, the Cornell Job Descriptive Index (JDI), devised by Smith and her associates (Hulin, Smith, Kendall, & Locke, 1963; Kendall, Smith, Hulin, & Locke, 1963; Locke, Smith, Hulin, & Kendall, 1963; Macaulay, Smith, Locke, Kendall, & Hulin, 1963; Smith, 1963; Smith & Kendall, 1963), and a modification of Porter's (1961, 1962, 1963a, 1963b, 1963c) job-satisfaction measure. The criteria developed by Campbell and Fiske (1959) for convergent and discriminant validity will be used to determine the strength of relationships between the measures.

METHOD

*Instruments*³

The Job Descriptive Index. This measure taps five areas of job satisfaction, which are presented in

¹ The data reported here were gathered while the author was a graduate student at Yale University. The author wishes to acknowledge the helpful comments provided by the following colleagues at Yale and elsewhere: Clay Alderfer, Chris Argyris, Lee Bolman, Vernon Buck, Tim Hall, Ed Lawler, and Lyman Porter. In addition, in two organizations, managers, supervisors, and workers gave willingly of their time: This is greatly appreciated.

² Requests for reprints should be sent to the author, School of Business, University of Toronto, 119 St. George Street, Toronto 5, Ontario, Canada.

³ Copies of both instruments (as used for the utility sample) have been deposited with the American Society for Information Science. Order NAPS

Table 1. It is a measure about which Vroom (1964) has commented:

[The JDI] is without doubt the most carefully constructed measure of job satisfaction in existence today. . . . The extensive methodological work underlying this measure as well as the available norms should insure its widespread use in both research and practice" [p. 100].

The format of this measure is quite simple; for each of the facets of job satisfaction, respondents are asked whether a series of words or phrases describes that particular aspect.

The goal-attainment measure. The original instrument (Porter, 1961, 1962, 1963a, 1963b, 1963c), based upon Maslow's (1954) hierarchy of needs, was used to tap the need satisfaction and need importance of a large sample (1,900) of American managers. It has the following form: For a series of need items, the respondent was asked to rate (on a 7-point scale): (a) How much is there now? (b) How much should there be? (c) How important is it to you? Need satisfaction is the difference between *b* and *a*, that is, a difference between levels of aspiration and attainment in each need area. Here, two modifications were made to the instrument. First, and at an operational level, the items were rewritten so as to be applicable to a sample of lower-level employees in organizations. The second change was of a conceptual nature. The question about level of aspiration was omitted so the question took on the characteristics of a goal-attainment and goal-importance measure;⁴ that is, the respondent was asked: (a) How much is there now? (b) How important is it to you? The goals tapped by this measure are presented in Table 1. The responses to the first question were taken to be measures of goal

Document 00196 from ASIS National Auxiliary Publications Service, c/o CCM Information Sciences, Inc., 22 West 34th Street, New York, New York 10001; remitting \$1.00 for microfiche or \$3.00 for photocopies.

⁴ See Evans (1969) for some discussion of the conceptual and operational confusion in the measurement of goal attainment and need satisfaction.

TABLE 1
FACETS OF JOB SATISFACTION AND GOAL IM-
PORTANCE TAPPED BY THE MEASURES

| Job Descriptive Index | Goal attainment |
|----------------------------------|---|
| Satisfaction with | Attainment of |
| Pay | Pay and fringe benefits |
| Work itself | Doing a good job |
| Opportunities for pro- motion | |
| Supervision | Respect from super- vision |
| Fellow workers | Respect from fellow workers |
| | Improving skills and abilities |
| | Job security |
| | Serving others (hospital only) |
| | Respect from doctors (hospital only) |

attainment. Each item was assumed to load onto one or more of the goals. A score for the attainment of each goal was obtained by summing across the appropriate items.⁵

The convergent and discriminant validation involved the JDI scores and the goal-attainment scores; in other words comparison was made between job satisfaction and goal attainment. These should

⁵ The items and their related goals are to be found in the NAPS materials.

be related, though not as strongly as might two different measures of satisfaction (Evans, 1969).

Samples

The questionnaires were administered as part of a larger study to two groups:

- 1. Workers in a public utility (n = 311). This organization will subsequently be referred to as "utility."
- 2. Nurses in a medium-sized general hospital (n = 88). This organization will subsequently be referred to as "hospital."

RESULTS

Convergent and Discriminant Validity

The two instruments do not have complete overlap in the aspects of job satisfaction or goal attainment that they purport to measure (see Table 1); this complicates the determination of validity. Tables 2 (utility) and 3 (hospital) present the complete intercorrelational matrices for the instruments. Campbell and Fiske (1957) have established four criteria for validity. Of the following criteria, No. 1 is for convergent validity; Nos. 2-4, discriminant validity.

- 1. Entries in the validity diagonal (circled, Tables 2 and 3) should be high and significantly different from zero. In other words,

TABLE 2
UTILITY: CORRELATIONS BETWEEN GOAL ATTAINMENT AND THE JOB DESCRIPTIVE INDEX

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | M | SD |
|------------------------------------|----|----|----|----|----|-----|---|---|---|----|----|----|----|----|-------|-------|
| Goal attainment | | | | | | | | | | | | | | | | |
| 1. Pay ^a | | | | | | | | | | | | | | | 4.86 | 1.26 |
| 2. Supervision ^a | 44 | | | | | | | | | | | | | | 4.93 | 1.54 |
| 3. Fellow workers ^a | 33 | 48 | | | | | | | | | | | | | 5.06 | 1.23 |
| 4. Work itself ^a | 44 | 67 | 71 | | | | | | | | | | | | 4.95 | 1.19 |
| 5. Skills & abilities ^a | 43 | 51 | 58 | 70 | | | | | | | | | | | 4.45 | 1.53 |
| 6. Security ^a | 35 | 20 | 20 | 26 | 24 | | | | | | | | | | 5.45 | 2.02 |
| 7. Promotion | | | | | | | | | | | | | | | — | — |
| JDI | | | | | | | | | | | | | | | | |
| 8. Pay ^b | 59 | 23 | 12 | 19 | 27 | 22 | | | | | | | | | 11.63 | 6.48 |
| 9. Supervision ^b | 21 | 60 | 18 | 34 | 28 | -00 | | | | | | | | | 37.89 | 12.83 |
| 10. Fellow workers ^b | 09 | 13 | 25 | 17 | 24 | -09 | | | | | | | | | 41.10 | 11.80 |
| 11. Work itself ^b | 15 | 23 | 22 | 29 | 38 | 04 | | | | | | | | | 32.68 | 10.60 |
| 12. Skills & abilities | | | | | | | | | | | | | | | — | — |
| 13. Security | | | | | | | | | | | | | | | — | — |
| 14. Promotion ^b | 31 | 26 | 35 | 36 | 51 | 12 | | | | | | | | | 9.93 | 7.62 |

Note.—Entries in the validity diagonal are circled. Broken line triangles are heterotrait-heteromethod triangles; solid line, heterotrait-monomethod triangles. n = 311.
^a Scales developed from the goal-attainment measure.
^b Scales in the JDI.

TABLE 3
HOSPITAL: CORRELATIONS BETWEEN GOAL-ATTAINMENT AND THE JOB DESCRIPTIVE INDEX

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | M | SD |
|------------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|---|----|----|----|----|----|----|----|----|----|-------|------|
| Goal attainment | | | | | | | | | | | | | | | | | | | | |
| 1. Pay ^a | 18 | | | | | | | | | | | | | | | | | | 4.61 | 1.00 |
| 2. Supervision ^a | 18 | 66 | | | | | | | | | | | | | | | | | 5.84 | 1.08 |
| 3. Fellow workers ^a | 20 | 77 | 67 | | | | | | | | | | | | | | | | 5.08 | 1.14 |
| 4. Work itself ^a | 32 | 32 | 46 | 38 | | | | | | | | | | | | | | | 5.51 | 1.04 |
| 5. Skills & abilities ^a | -04 | -12 | -01 | 05 | 07 | | | | | | | | | | | | | | 5.05 | 1.18 |
| 6. Security ^a | 48 | 32 | 41 | 41 | 64 | 03 | | | | | | | | | | | | | 5.71 | 1.60 |
| 7. Serve others ^a | 19 | 59 | 76 | 59 | 47 | 05 | 55 | | | | | | | | | | | | 5.97 | 0.92 |
| 8. Doctors ^a | | | | | | | | | | | | | | | | | | | 5.65 | 1.20 |
| 9. Promotion | | | | | | | | | | | | | | | | | | | — | — |
| JDI | | | | | | | | | | | | | | | | | | | | |
| 10. Pay ^b | 46 | 08 | -09 | -05 | 12 | 05 | 12 | -10 | | | | | | | | | | | 9.21 | 5.60 |
| 11. Supervision ^b | 05 | 31 | 16 | 25 | -03 | -04 | -02 | 13 | | | | | | | | | | | 42.85 | 9.04 |
| 12. Fellow workers ^b | 02 | 10 | 23 | 03 | 16 | 28 | 21 | 26 | | | | | | | | | | | 45.51 | 8.25 |
| 13. Work itself ^b | 39 | 01 | 15 | 16 | 29 | 14 | 40 | 17 | | | | | | | | | | | 34.66 | 7.10 |
| 14. Skills & abilities | | | | | | | | | | | | | | | | | | | — | — |
| 15. Security | | | | | | | | | | | | | | | | | | | — | — |
| 16. Serve others | | | | | | | | | | | | | | | | | | | — | — |
| 17. Doctors | | | | | | | | | | | | | | | | | | | — | — |
| 18. Promotion ^b | 21 | 19 | 16 | 07 | 46 | 06 | 24 | 25 | | | | | | | | | | | 9.35 | 6.75 |

Note.—Entries in the validity diagonal are circled. Broken line triangles are heterotrait-heteromethod triangles; solid line, heterotrait-monomethod triangles. $n = 88$.
^a Scales developed from the goal-attainment measure.
^b Scales in the JDI.

correlations between the two instruments on the same variables should be high. Utility: requirement met. Hospital: requirement met, though that for the work scale is low ($r = .16$, $p < .05$).

2. A validity diagonal value should be higher than values lying in its column and row in the heterotrait-heteromethod triangles (broken line triangles). As an example, the correlation between pay and pay measured with each instrument should be higher than the correlation between pay measured with one instrument and any other variable measured with the *other* instrument. A simple sign test was used to evaluate differences (Siegel, 1956, pp. 68-74). Utility: requirement met for pay ($p = .002$), supervision ($p = .002$), and fellow workers ($p = .02$), but not for work itself ($p = .254$). Hospital: requirement met for pay ($p = .001$), supervision ($p = .001$), and fellow workers ($p = .033$), but not for work itself ($p = .113$).

3. A validity diagonal value should be higher than the corresponding values in the heterotrait-monomethod triangles (solid line triangles). For example, the correlation between pay and pay measured with each instrument should be higher than the correlation between pay measured with one instrument and any other variable measured with the same instrument. Again, a simple sign test was used to evaluate differences.⁶ Utility: the requirement was met for pay ($p = .031$), and to some extent for supervision ($p = .188$) where the only heterotrait-monomethod correlation exceeding the validity diagonal value is that for supervision-work. Hospital: the requirement was met for pay ($p = .003$), but not for the other scales.

4. There should be shown the same pattern of trait interrelationships in all the heterotrait triangles of both the monomethod and heteromethod blocks. This is essentially a question of the ordering of the correlation coefficients within each block. To test the degree of agreement between the orderings in each block, the correlation coefficients were ranked by size, and the coefficient of concor-

dance, W (Siegel, 1956, pp. 229-238), was computed.⁷ Utility: this requirement was met, $W = .59$ ($p = .05$). Hospital: this requirement was met, $W = .57$ ($p = .05$).

CONCLUSION

The results are encouraging. The criteria proposed by Campbell and Fiske (1959) are very rigorous. Few of the studies they reported met all four of their criteria. For the JDI and the goal-attainment measure, all four criteria were met in at least one of the samples. It is highly likely that they are valid. The use of the coefficient of concordance to measure the fourth criterion is an advance on the examination of the matrix by eye that was proposed by Campbell and Fiske (1959).

⁷ The fourth stage of the check was made on the reduced matrix for which both instruments tap common dimensions, that is, Rows 1-4/Columns 1-4, Rows 10-13/Columns 1-4, Rows 10-13/Columns 10-13, for the hospital; and Rows 1-4/Columns 1-4, Rows 8-11/Columns 1-4, Rows 8-11/Columns 8-11, for the utility.

REFERENCES

- CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, 56, 81-105.
- EVANS, M. G. Conceptual and operational problems in the measurement of job satisfaction. *Journal of Applied Psychology*, 1969, 53, 93-101.
- HULIN, C. L., SMITH, P. C., KENDALL, L. M., & LOCKE, E. A. *Cornell studies of job satisfaction: II. Model and method of measuring job satisfaction*. Ithaca: Cornell University, 1963. (Mimeo)
- KENDALL, L. M., SMITH, P. C., HULIN, C. L., & LOCKE, E. A. *Cornell studies of job satisfaction: IV. The relative validity of the Job Descriptive Index and other methods of measurement of job satisfaction*. Ithaca: Cornell University, 1963. (Mimeo)
- LOCKE, E. A., SMITH, P. C., HULIN, C. L., & KENDALL, L. M. *Cornell studies of job satisfaction: V. Scale characteristics of the Job Descriptive Index*. Ithaca: Cornell University, 1963. (Mimeo)
- MACAULAY, D. A., SMITH, P. C., LOCKE, E. A., KENDALL, L. M., & HULIN, C. L. *Cornell studies of job satisfaction: III. Convergent and discriminant validity for measures of job satisfaction by rating scales*. Ithaca: Cornell University, 1963. (Mimeo)
- MASLOW, A. H. *Motivation and personality*. New York: Harper, 1954.

⁶ The second and third stages of the convergent and discriminant validity check were made on the complete correlation matrix.

- PORTER, L. W. A study of perceived need satisfactions in bottom and middle management jobs. *Journal of Applied Psychology*, 1961, **45**, 1-10.
- PORTER, L. W. Job attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology*, 1962, **46**, 375-384.
- PORTER, L. W. Job attitudes in management: II. Perceived importance of needs as a function of job level. *Journal of Applied Psychology*, 1963, **47**, 141-148. (a)
- PORTER, L. W. Job attitudes in management: III. Perceived deficiencies in need fulfillment as a function of line versus staff type of job. *Journal of Applied Psychology*, 1963, **47**, 267-275. (b)
- PORTER, L. W. Job attitudes in management: IV. Perceived deficiencies in need fulfillment as a function of size of company. *Journal of Applied Psychology*, 1963, **47**, 386-397. (c)
- SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.
- SMITH, P. C. *Cornell studies of job satisfaction: I. Strategy for the development of a general theory of job satisfaction*. Ithaca: Cornell University, 1963. (Mimeo)
- SMITH, P. C., & KENDALL, L. M. *Cornell studies of job satisfaction: VI. Implications for the future*. Ithaca: Cornell University, 1963. (Mimeo)
- VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

(Received March 7, 1968)

Manuscripts Accepted for Publication in the

Journal of Applied Psychology

- Examination of Some Measures of Creative Ability by the Multitrait-Multimethod Matrix: P. Goodman,* J. Furcon, and J. Rose: Graduate School of Business, University of Chicago, 5836 Greenwood Avenue, Chicago, Illinois 60637.
- Perceived Need Satisfactions of Domestic versus Overseas Managers: John M. Ivancevich*: Department of Administrative Sciences and Quantitative Methods, Commerce Building, University of Kentucky, Lexington, Kentucky 40506.
- Use of the Work Components Study (WCS) with New College-Level Employees: R. N. Ford, E. F. Borgatta,* and G. W. Bohrnstedt: Social Science Building, University of Wisconsin, Madison, Wisconsin 53706.
- Validation Study of Polygraph Examiner Judgments: Philip J. Bersh*: Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.
- Feedback and Response Mode in Performing a Bayesian Decision Task: David W. Martin* and Charles F. Gettys: Human Performance Center, Ohio State University, 404-B West 17th Avenue, Columbus, Ohio 43210.

* Asterisk indicates author for whom address is supplied.

A STUDY OF SOME EPPS VARIABLES AS FACTORS OF ACADEMIC ACHIEVEMENT

R. P. BHATNAGAR ¹

Regional College of Education, Ajmer, India

The study related six EPPS personality variables to the academic achievement of 261 high school students after controlling the effects of socioeconomic status, intelligence, school differences, and age differences. It was found that the relationship between personality and achievement is tied with age levels, intelligence, and specificity of academic achievement. The research concluded that the EPPS (Hindi) variables do contribute to academic achievement, but differentially at different levels of age and intelligence and for different types of academic achievement such as arts and science achievement.

Various aspects of personality such as motivation, maturation, socialization, emotional adjustment and self-fulfillment, self-concept, ego-organization, identification processes, etc., have been studied as factors of academic achievement. A great volume of psychological theory has accumulated around personality factors of academic achievement. A variety of methods has been used in research studies of this kind. Mostly, case-study methods and group-study methods involving the use of psychological tests have been employed. Ephron (1953), Mehus (1953), Gann (1945), Fernald (1943), Kimball (1953), and Conklin (1940) have used case-study methods for studying factors of academic failure. All of them have found that inadequacy of personality in some form is an essential characteristic of the children who are academically unsuccessful. But the results of research studies have not been definite. Rather, as was held by Gowan (1960, p. 91); "The problem appears more complex than it was first indicated." This situation necessitates additional replications of such studies, especially cross-cultural studies. The present study is an attempt in this direction.

¹ The author is highly indebted to P. E. Vernon of the University of London, Institute of Education, for the valuable suggestions made on request about this piece of research in his letter of May 29, 1963, and to Helen M. Walker of Columbia University for her valuable guidance regarding statistical analysis and treatment of the data provided when she was in India as consultant in the National Council of Educational Research and Training, Delhi. Requests for reprints should be sent to the author, Reader in Education, Regional College of Education, Ajmer India.

The problem in the present study was to investigate the relationship that may exist between certain personality variables and the academic achievement of high school students. The personality variables selected for this purpose were the personality needs as measured by the EPPS developed in Hindi by the investigator (Bhatnagar, 1966a). All the 15 needs have not been used. Only 6 needs which were found to differentiate between the two groups of underachievers and overachievers were considered valid for the purpose (Bhatnagar, 1966b). They are need Achievement, need Dominance, need Autonomy, need Nurturance, need Endurance, and need Aggression.

METHOD

Subjects

The population in this study has been defined as the male students of Class IX of the city of Moradabad, U.P., India. The population was listed in the form of intact schools. There were 12 schools in all having about 1,000 students in Class IX.

From the above population, a sample of six schools having 612 students was randomly drawn. All the 612 students were not included in the final analysis. Some students had to be excluded as they did not take all the tests, and records of age, etc., were not available for them. Control of the socioeconomic status variable through cross-tabulation also reduced the size of the sample. Finally, 261 students were available for the purpose of analysis.

Data

The data were collected for all the 612 students. These included (a) ages, (b) school marks in the form of composite *T* scores over six subjects, (c) scores on the Central Institute of Education (1959) test of intelligence, (d) scores on six personality

needs measured by the EPPS (Hindi), and (e) scores on a socioeconomic status questionnaire.

The scores on these six need scales were correlated with achievement scores after eliminating the effects of intelligence and school differences. The results were analyzed separately for each of the three age groups into which the total sample was split.

Design

An attempt was made to control five factors: socioeconomic status, age, intelligence, school differences, and sex. Failure to control these factors would have caused a distortion of relationship that might exist between personality needs and academic achievement. The control of these factors was achieved in different ways. The sex bias was eliminated by confining the experiment to the population of boys only. Age effects were minimized by splitting the sample covering a total age range of 6 yr. into three subgroups, each with a 2-yr. age range, and analyzing the results separately for each age group. Socioeconomic status was controlled through cross-tabulation by eliminating the top and the bottom extreme groups from the distribution and using only the middle group which may be considered roughly homogeneous with respect to this variable. School differences were eliminated by computing correlations between personality variables and academic achievement by using the within-school sums of squares and cross-products and applying the formula given by Garrett (1958). The within-schools correlations are considered unaffected by differences in school means. Intelligence was controlled by computing partial correlations (intelligence being partialled out) between personality variables and academic achievement.

RESULTS

The results are presented in the following sections. The first part contains a discussion of the relationship between personality and academic achievement in general, while the second part considers the relationship when specific achievements in arts and science are treated separately.

General Achievement

When age, intelligence, and the specificity of achievement are all disregarded (Table 1, total correlations), the nurturance and endurance needs are positively related and dominance is negatively related to achievement. When intelligence is partialled out and other factors are allowed to vary (Table 1, partial correlations), only nurturance and endurance needs remain significantly related. The dominance need fails to hold a significant relationship. Table 2 shows correlations between

TABLE 1
CORRELATIONS BETWEEN PERSONALITY VARIABLES
AND ACHIEVEMENT: AGE AND SPECIFICITY
OF ACHIEVEMENT DISREGARDED

| Variable | Total correlations | Partial correlations |
|----------|--------------------|----------------------|
| n Ach | .056 | .028 |
| n Aff | -.087 | -.017 |
| n Dom | -.128* | -.032 |
| n Nur | .171** | .149** |
| n End | .202** | .145* |
| n Agg | -.027 | .025 |

* $p < .05$.
** $p < .01$.

EPPS (Hindi) variables and achievement (general) when intelligence is not partialled out. Table 3 presents partial correlations between the same set of variables when intelligence is partialled out.

The partialled correlations can be considered comparatively least affected by the correlated variables. The nurturance ($r = .332$, $p < .01$; $r = .421$, $p < .01$) and endurance ($r = .235$, $p < .05$; $r = .221$, $p < .01$) needs are positively related to general achievement in the first (15.5–17.5 yr.) and second (13.5–15.5 yr.) age groups when intelligence is partialled out (Table 3). But in the third group of the youngest students, approximately 12.5 yr. old, they are not related. The need for achievement is positively related to academic achievement for Group II when intelli-

TABLE 2
CORRELATIONS BETWEEN ACADEMIC ACHIEVEMENT
(GENERAL) AND PERSONALITY VARIABLES:
INTELLIGENCE NOT PARTIALED OUT

| Variables | Group I (15.5–17.5 yr.) ^a | Group II (13.5–15.5 yr.) ^b | Group III (11.5–13.5 yr.) ^c | Total ^d |
|-----------|---|--|---|--------------------|
| n Ach | .095 | .199* | -.173 | .056 |
| n Aff | -.90 | -.064 | .250* | -.087 |
| n Dom | .135 | .016 | -.130 | -.128* |
| n Nur | .201 | .370** | .180 | .171** |
| n End | .260* | .239* | .130 | .202** |
| n Agg | .036 | .197* | .150 | -.027 |

^a $N = 76$.
^b $N = 120$.
^c $N = 65$.
^d $N = 261$.
* $p < .05$ (two-tailed).
** $p < .01$ (two-tailed).

TABLE 3

PARTIAL CORRELATIONS BETWEEN ACADEMIC
ACHIEVEMENT (GENERAL) AND PERSON-
ALITY VARIABLES: INTELLIGENCE
PARTIALED OUT

| Variable | Group I (15.5– 17.5 yr.) ^a | Group II (13.5– 15.5 yr.) ^b | Group III (11.5– 13.5 yr.) ^c | Total ^d |
|----------|---|--|---|--------------------|
| n Ach | .127 | .150 | –.256* | .028 |
| n Aff | –.051 | –.063 | .255* | –.017 |
| n Dom | .153 | .198* | –.060 | –.032 |
| n Nur | .332** | .421** | .151 | .149* |
| N End | .235* | .221* | .095 | .145* |
| n Agg | .060 | .035 | .121 | .025 |

^a $N = 76$.

^b $N = 120$.

^c $N = 65$.

^d $N = 261$.

* $p < .05$ (two-tailed).

** $p < .01$ (two-tailed).

gence is not held constant (Table 2), but loses its significance for that group and becomes negatively related for the third group when intelligence is partialled out (Table 3). Such changes in the position of significant needs over the age groups and within the age groups when intelligence is partialled out are noticed in the cases of need Affiliation and need Aggression also (compare Tables 2 and 3). This could imply that the variability in intelligence affects the relationship that might exist between personality and achievement. It seems that the relationship between personality factors and academic achievement is not of a general nature in the sense that it is the same at all levels of ability. This could mean that the relationship of personality with academic achievement is tied with intelligence.

It appears from Table 3 that EPPS needs contributing to academic success differ from one age group to another. In the first age group (15.5–17.5 yr.), it is the configuration of positively related nurturance ($r = .332$, $p < .01$) and endurance ($r = .235$, $p < .05$); in the second (13.5–15.5 yr.), of dominance ($r = .198$, $p < .05$), nurturance ($r = .421$, $p < .01$), and endurance ($r = .221$, $p < .05$); and in the third (11.5–13.5 yr.), of affiliation ($r = .255$, $p < .05$) and negatively related need Achievement ($r = -.256$, $p < .05$). This suggests that the status of EPPS (Hindi) needs as predictors of academic achievement is determined by the age level also. It appears

that the relationship of personality with academic achievement is, probably, not the same at all age levels.

The overall picture of the three age groups suggests that almost all the six EPPS (Hindi) needs except achievement in the third group are positively correlated with academic achievement. It may be concluded that personality factors do contribute to academic achievement at the high school level. However, it is surprising that the achievement need is not related to academic achievement in the first and the second age groups (range, from 13.5 to 17.5 yr.) and emerges with a significantly negative correlation in the third age group of students, approximately 12.5 yr. old. A nonsignificant relationship between need Achievement and academic performance has been found in a few studies (Lowell, 1950; McClelland, Atkinson, Clark, & Lowell, 1953), but a negative relationship appears to be unprecedented.

Specific Achievement

On certain grounds it was hypothesized that the relationship between personality factors and academic achievement might not be uniform for different types of achievement. It was thought that the findings obtained with reference to achievement in general as discussed above might not hold in a situation in which success in arts or science is predicted separately. For this reason each age group was further broken down into two subgroups, the arts group and the science group. For each of these groups partial correlations between EPPS variables and specific achievements in arts and science (intelligence having been partialled out) were computed. These correlations are shown in Table 4.

It is observed that the direction and magnitude of relationship of EPPS needs with specific achievement vary from arts to science group at each age level.

In the first group need Endurance emerges as a predictor of achievement in arts, while need Achievement predicts achievement in science. In the second age group achievement in arts is unrelated to any of the six personality variables, while achievement in science is related to two needs, need Achievement and

TABLE 4
PARTIAL CORRELATIONS BETWEEN PERSONALITY VARIABLES AND ACADEMIC ACHIEVEMENT
FOR ARTS AND SCIENCE STUDENTS OF DIFFERENT AGE GROUPS

| Variable | Group I | | | Group II | | | Group III | | |
|----------|---------|---------|---------|----------|---------|---------|-----------|---------|---------|
| | Arts | Science | General | Arts | Science | General | Arts | Science | General |
| n Ach | .106 | .375* | .127 | .165 | -.240 | .150 | -.220 | -.221 | -.256* |
| n Aff | -.229 | -.259 | -.051 | -.115 | -.571** | -.063 | .020 | .222 | .255* |
| n Dom | .213 | -.200 | .153 | -.150 | -.129 | .198* | .070 | -.180 | -.060 |
| n Nur | .314* | .210 | .332 | -.090 | .250* | .421** | .092 | -.331* | .151 |
| n End | .106 | .210 | .235* | .090 | .166 | .221* | .389* | .027 | .095 |
| n Agg | -.022 | .050 | .060 | -.285* | -.040 | .035 | -.386* | .378* | .121 |

* $p < .05$ (two-tailed).
** $p < .01$ (two-tailed).

Affiliation. In the third age group need Aggression is negatively related to achievement in arts and positively related to achievement in science. Additionally, for the same age group endurance is positively related to arts achievement and unrelated to science achievement. Similarly, for the same age group need Nurture is negatively related to science achievement, but unrelated to arts achievement. This suggests that, probably, specificity of achievement also affects the status of EPPS needs as predictors of academic achievement.

Personality needs are the promising predictors of academic achievement in arts and science courses at different age levels.

In the case of the 16.5-yr.-olds need Nurture is significantly correlated with achievement in arts, the coefficient of correlation being .314. This implies that probably 9% of the success variance in arts is dependent on this need. Achievement in science for the same age group is predictable from need Achievement to the extent of 16%, the coefficient of correlation being .375. In the second group of those approximately 14.5 yr. old none of the six EPPS variables is a significant predictor of achievement in arts. Achievement in science, however, is significantly correlated with the achievement and affiliation needs. But, both the correlations are negative. In the third age group of those approximately 12.5 yr. old the endurance need appears to predict arts achievement in the positive direction to the extent of 16%, the corre-

lation being .389. The aggression need does so in the negative direction ($r = -.386$). The science achievement in this group is positively predictable from need Aggression ($r = .378$) and negatively predictable from need Nurture ($r = -.331$).

DISCUSSION

The contribution of EPPS variables to academic achievement is presented separately.

Need for Achievement

The achievement need fails to predict achievement in general for the first and the second groups covering students between the ages of 13.5 and 17.5 yr. In the third group of those approximately 12.5 yr. old it is negatively related to school success. The achievement motivation fails to predict success in arts at all age levels. It predicts success in science for the first group of approximately 16.5-yr.-old students. For the other age groups it fails as a predictor.

Need for Affiliation

The affiliation motive is positively related to general achievement in the third age group of 12.5-yr.-old students. In the case of specific achievement in science it is negatively related for the second age group. It does not predict arts achievement at any age level.

Need for Dominance

The dominance need is related to general achievement only for the second age group.

Need for Nurturance

The nurturance need emerges as a positive predictor of general achievement for the first and the second group. For the third group it fails to predict general achievement. It is positively related ($r = .314$, $p < .05$) to arts achievement for the first age group (15.5–17.5 yr.). At other age levels it is not related to achievement in arts. It is negatively related ($r = -.331$, $p < .05$) to science achievement for the third group and positively related for the second group. It does not predict achievement in science for the first age group.

Need for Endurance

The endurance need is related to general achievement in the first (15.5–17.5 yr.) and the second (13.5–15.5 yr.) age groups, but not to the third age group. It does not predict science achievement for any age group. In the case of achievement in arts it is positively related ($r = .389$, $p < .05$) only to the third group (11.5–13.5 yr.). In other age groups it fails to predict achievement in arts.

Need for Aggression

The aggression need is not related to achievement in general at any age level. With achievement in arts it emerges as negatively related ($r = -.285$, $-.386$, respectively) to the second and third age groups. It does not, however, predict arts achievement for the first group. With achievement in science in the third age group it correlates positively ($r = .378$, $p < .05$). In other age groups it does not correlate with science achievement.

REFERENCES

- BHATNAGAR, R. P. Development of the Edwards Personal Preference Schedule in Hindi. *Indian Psychological Review*, 1966, 2, 103–108. (a)
- BHATNAGAR, R. P. Personality needs and achievement. (Regional College of Education, Ajmer) *The Alumnus*, August 1966. (b)
- CENTRAL INSTITUTE OF EDUCATION. *Verbal group test of intelligence manual of instructions*. (Publ. No. 13) Delhi, India: Central Institute of Education, 1959.
- CONKLIN, A. M. Failure of highly intelligent pupils. Columbia University Teachers College of New York: *Contributions to Education*, 1940, No. 792.
- EPHRON, B. K. *Emotional difficulties in reading*. New York: Julian Press, 1953.
- FERNALD, G. M. *Remedial techniques in basic school subjects*. New York: McGraw-Hill, 1943.
- GANN, E. *Reading difficulty and personality organization*. New York: Kings Crown Press, 1945.
- GARRETT, H. E. *Statistics in psychology and education*. New York: Longmans, Green, 1958.
- GOODSTEIN, L. D., & HEILBRUN, A. B. Prediction of college achievement from the Edwards Personal Preference Schedule at three levels of intellectual ability. *Journal of Applied Psychology*, 1962, 46, 317–320.
- GOWAN, J. C. Factors of achievement in high school and college. *Journal of Counseling Psychology*, 1960, 7, 91–95.
- KIMBALL, B. Case studies in educational failure during adolescence. *American Journal of Orthopsychiatry*, 1953, 23, 406–415.
- LOWELL, E. L. A methodological study of projectively measured achievement motivation. Unpublished master's thesis, Wesleyan University, 1950.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
- MEHUS, H. Learning and therapy. *American Journal of Orthopsychiatry*, 1953, 23, 416–421.

(Received March 9, 1968)

EXPERIENCE AND PRIOR PROBABILITY IN A COMPLEX DECISION TASK ¹

MICHAEL H. STRUB ²

Human Performance Center, Ohio State University

Six experienced and six naive Ss evaluated probabilistic data, determined sources of data generation, and predicted subsequent data in a complex decision task. Experience and prior probability were combined factorially. Results indicated that experienced Ss (*a*) were less conservative data evaluators, (*b*) determined data sources on the basis of fewer data samples, (*c*) were more sensitive to prior-probability values, and (*d*) adopted a maximization strategy in prediction more consistently than did naive Ss. The importance of using trained personnel in the evaluation of realistic decision capabilities and the need for caution in generalizing from data obtained from naive Ss who serve in most laboratory studies of decision making were discussed.

The process of decision making usually involves evaluation of data from the environment and subsequent selection of action with respect to the environment. Evaluation of data can be defined as a determination of the extent to which a piece or pieces of evidence (data) favor the truth of one state of the environment over others, while action selection is a response made after a state of the environment has been judged true by the decision maker. For example, a falling barometer and high humidity are data predictive of rain. The carrying of an umbrella by the decision maker indicates that he has judged rain to be the projected state of the environment, perhaps, in part, because he has observed these data.

It has often been found in laboratory situations that decision makers are conservative data evaluators (for a discussion of "conservatism" see Peterson & Beach, 1967). In

addition, they tend to purchase more data than is formally optimal prior to action selection (Irwin & Smith, 1957; Lanzetta & Kanareff, 1962; Swets & Green, 1961). However, nonlaboratory decision makers, such as those involved in military command-control systems, are considerably more experienced than Ss who serve in most laboratory studies of decision making. In view of this fact, it becomes meaningful to ask whether *experienced* decision makers also display conservatism in evaluating and purchasing data. According to Edwards (1966), evidence bearing upon this issue is sparse. One purpose of the present study, therefore, was to investigate the importance of *experience* in decision making.

Another characteristic of real-world decision situations is the presence of a data history. In many laboratory simulations of decision making (e.g., Kaplan & Newman, 1966), *S* begins with complete ignorance of the environment: He does not regard any state as more likely than any other, and he is void of any experience with data upon which to base a prediction of the true state. In reality, people are rarely so poorly equipped when they enter the decision situation. Prior knowledge that there is a predominance of rainy over clear weather in a given location, for example, may well be combined with the barometer and humidity check in deciding to carry an umbrella. Assuming that the data to be processed do not conflict with prior knowledge (e.g., the barometer and humidity

¹ The research reported in this paper was carried out at the Human Performance Center and was supported by the Aeromedical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, under Contract No. AF 33(615)-2248 with the Ohio State University Research Foundation. Further reproduction is authorized to satisfy the needs of the United States Government. An expanded version of this study was submitted by the author in partial fulfillment of the requirements for the master's degree at Ohio State University.

² Requests for reprints should be sent to the author, Human Performance Center, Ohio State University, 404-B West 17th Avenue, Columbus, Ohio 43210.

data do not favor sunshine), incorporation of prior-probability information into the aggregate prediction should encourage action to be taken after collection of fewer diagnostic data than would otherwise be the case. Such an effect was suggested by the descriptive data of Green, Halbert, and Minas (1964), but failed to appear in a direct investigation of prior probability and data purchase reported by Messick (1964). One reconciling explanation is that Ss had more experience in the type of task employed by Green et al. Experience may be a moderator variable which increases the decision maker's sensitivity to prior probability. Another objective of the present experiment, then, was to test this hypothesis directly by combining prior probability and experience factorially within the same design.

The familiar ball and urn task which is commonly used to assess data evaluation by human decision makers requires *S* to decide, on the basis of samples of balls drawn, which of two urns is being sampled: one containing predominantly red balls or one containing predominantly black balls. This task becomes a probability-learning task if *S* is instructed actually to *predict* the color of the next draw. Correct response maximization in probability learning is achieved by the exclusive prediction of the more frequent datum, a strategy which Ss rarely adopt except under conditions of monetary payoff and extended training (Edwards, 1961). If Ss were to make such predictions while concurrently *evaluating* the data, they might well be more prone to exploit the maximization rule, especially if they were experienced in data evaluation. A final objective of the present research, then, was to discover if prediction behavior is influenced by prior and concurrent experience in data evaluation.

METHOD

Subjects

Six naive and six experienced male college students served as Ss. The latter had all served previously as decision makers in an ongoing simulation project devoted to the study of probabilistic information processing in command-control systems (see Southard, Schum, & Briggs, 1964). In conjunction with this program they received approximately 114 hr. of lecture sessions, demonstrations, problem-

solving sessions, and on-the-job training in an effort to maximize their proficiency in dealing with probabilistic information. In addition, these Ss had participated in a variety of probability-estimation experiments over a period of at least 3 mo. prior to this study. The naive Ss, selected for the ongoing simulation project on the basis of the same criteria as the experienced Ss, had received none of this formal training and had served in only one previous probability-estimation study of less than 1-wk. duration.

The experienced Ss were paid a base rate of \$1.50 to \$1.85 an hour according to their seniority in the project mentioned above, while all naive Ss received a base rate of \$1.25 an hour. It should be noted that in order to obtain realistically *experienced* Ss it was necessary to accept a degree of experimental confounding. Such Ss, by virtue of their experience, were (a) employed on a more permanent basis than the naive ones and (b) paid at a higher rate than could be justified for naive ones.

In addition to base pay, all Ss were rewarded or penalized in accordance with their performance on the two primary task requirements: action selection and prediction. For action selection, an optional stopping scheme was used in which a correct choice was rewarded by \$.10 minus \$.001 for each datum drawn (red or black marble). An incorrect choice was penalized by \$.10 plus \$.001 for each datum drawn. For the prediction task, each correct anticipation of the next datum was rewarded by \$.001 with no penalty for incorrect prediction.

Procedure

The following instructions acquainted Ss with the task characteristics:

I have two bowls, each of which contains a total of 100 red and black marbles. I am going to select a bowl and begin drawing from it, one at a time. After each draw, the marble is put back in the bowl. Before I begin to draw, you will be told the composition of each of the two bowls such as 60/40, 70/30. The first number of each composition is always with reference to the number of red marbles, and the second, the number of black. You will also be told the prior probability of selecting one or the other bowl.

Sampling was then initiated, and *S* indicated, after each draw (trial), which bowl (proportion) he thought was favored by the current accumulation of data and how confident he was in his evaluation. Both responses were made by marking response sheets: 1 and 2 were used to designate bowls, and probability values (.50-.99) were used to indicate confidence. Following this evaluation, *S* was required to indicate whether the next marble would be red (R) or black (B), again by marking his response sheet. At any trial in each series of 100 trials, *S* could make his single *action* decision by selecting 1 or 2 as the true source of data (bowl) and accepting the consequences (payoff or loss). The *E* re-

corded the trial number on which *S* made his action selection, but *S* continued to evaluate and predict data through all 100 samples. By the end of sampling there was usually little doubt as to the identity of the correct state (bowl), although no feedback was given directly until after the experiment.

Apparatus

The drawing of red and black marbles was simulated by an IBM 1401 computer programmed to generate random sequences according to the specified prior probability (50-50 or 90-10) and the bowl compositions (marble proportions of 60/40-70/30 or 80/20-90/10). Data were presented to *Ss* via a closed-circuit TV system. As each sample was presented, it registered on one of two adjacent counters which kept cumulative totals of each sample (thus providing a realistic "history" upon which to base decisions).

Design

The design consisted of a between-*S* variable (*experience*) and two within-*S* variables (*prior probability* and *bowl proportion set*), each administered at two levels. Each prior-probability and proportion-set combination occurred in 12 sequences; thus, a total of 48 sequences was viewed by each group. These were administered in 10 experimental sessions, each requiring about 4 hr. Since sequences within each combination varied according to a random-sampling procedure, any learning effects were confounded with the relative difficulty of the sequence. Thus, all trial and trial-interaction effects were collapsed into a single error estimate in the analyses. Dependent variables analyzed included final probability estimates (after 100 samples), number of samples purchased, and degree of maximization in prediction (i.e., the proportion of trials on which the more frequent event was predicted).

RESULTS AND DISCUSSION

Final Probability Estimates

Edwards (1966), in a discussion of the parametric difficulties involved in analyzing subjective probability scores, made a strong case for transformation of these scores to log-odds or log-likelihood ratios. Accordingly, *Ss*' final subjective probability values were transformed here to log-odds form, where $\log \text{ odds} = \log S's\ p - \log (1 - S's\ p)$. Figure 1 represents the mean log odds for groups obtained across prior-probability and proportion-set conditions.

An analysis of variance performed on these data indicated that the overall final subjective probability was significantly higher for the experienced than for the naive group, $F =$

8.68, $df = 1/10$, $p < .025$, and that *Ss* expressed significantly more certainty under 90-10 than 50-50 prior probability, $F = 10.61$, $df = 1/10$, $p < .01$. Proportion set was also significant, $F = 56.25$, $df = 1/10$, $p < .001$, suggesting that *Ss* found the data samples in the 80/20-90/10 set to be more diagnostic than those in the 60/40-70/30 set. However, the main effects of prior probability and proportion set must be interpreted in light of the Prior Probability \times Proportion Set interaction which also achieved significance, $F = 8.25$, $df = 1/10$, $p < .025$. The interaction is apparent in Figure 1: The increase in certainty across prior probability is restricted to the 60/40-70/30 proportion-set conditions.

The finding that the experienced *Ss*' final subjective probabilities were consistently higher than the naive *Ss*' suggests that the experienced *Ss* were less conservative than the naive *Ss*. However, any comparison of the present phenomenon with previous studies of

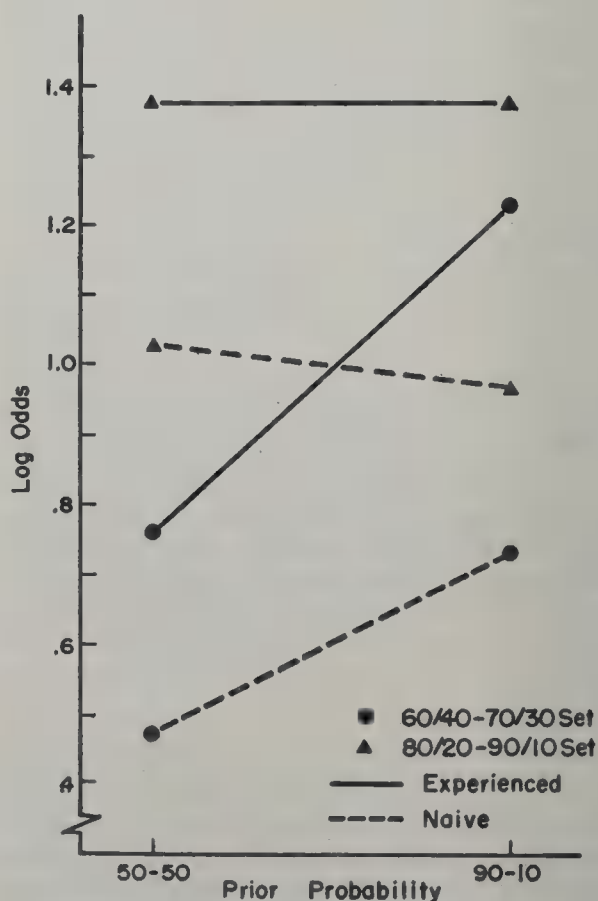


FIG. 1. Final subjective probability values transformed to log odds for experienced and naive *Ss* across prior probability and proportion set.

conservatism must be made cautiously. The present study focused on *final* subjective probability estimates, while conservatism is more commonly inferred from a comparison of the amount of datum-to-datum *revision* produced by Ss with that produced by Bayes' theorem. The only comparative estimates obtained in the present study were those involving the average of final probability estimates provided by experienced Ss, by naive Ss, and by the Bayes model. The mean of these final certainty estimates was .95 for Bayesian revisions, .94 for the experienced group, and .86 for the naive group. While such a comparison suggests that the experienced group was nearly "Bayesian" in its probabilistic estimates, it should be recognized that conclusions based upon average subjective probabilities are tenuous at best (Edwards, 1966).

The finding that Ss were more confident about the 80/20-90/10 than the 60/40-70/30 proportion set was accompanied by higher final Bayesian revisions within the former set, indicating that the randomly generated sequences were more diagnostic in the 80/20-90/10 set. The Prior Probability \times Proportion Set interaction suggests that in the more diagnostic proportion set the impact of the data was sufficient to eliminate the initial diagnostic value of the 90-10 prior probability, while in the less diagnostic proportion set prior probability still influenced probability estimates after 100 samples.

Number of Samples Purchased

Figure 2 indicates the mean number of samples purchased for each group across prior probability and proportion set. The results of an analysis of variance of these data yielded an experience effect which approached significance, $F = 4.32$, $df = 1/10$, $p < .07$, and a significant main effect of prior probability, $F = 5.06$, $df = 1/10$, $p < .05$. An F_{\max} test of the total variability within each group revealed heterogeneity of variance, $F = 1.57$, $df = 288$, $p < .01$. It was decided that the extensive intergroup variability warranted an analysis of the data by group means and by separate within-group analyses. Group means were compared by an exact randomization

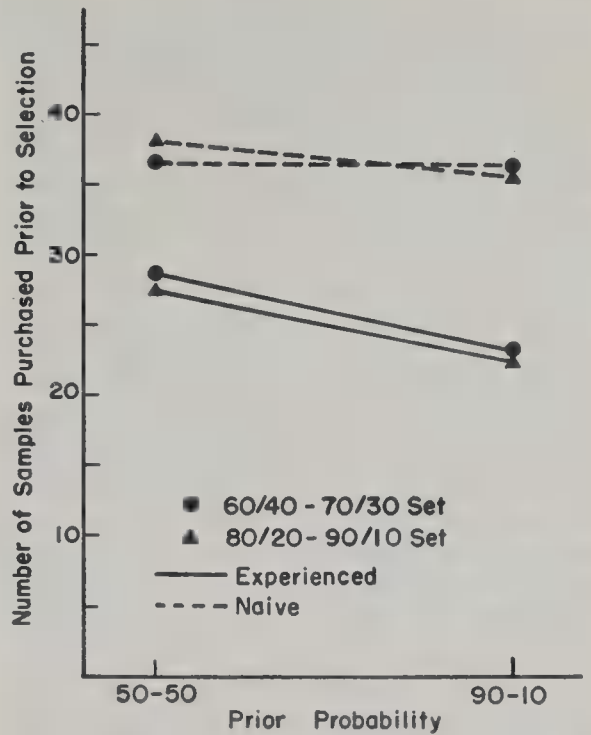


FIG. 2. Mean number of samples purchased for experienced and naive Ss across prior probability and proportion set.

test (Kempthorne, 1955; McHugh, 1963). For each level of prior probability and distribution set, the experienced group purchased fewer samples than the naive group. This outcome arrangement resulted in a significance level of .0625 which, based on the probability array of 2^4 or 16 outcomes, is the minimum value obtainable in the test (since $1/16 = .0625$).

In an analysis of variance performed on the number of samples purchased by the naive Ss alone, obtained F values for all effects were < 1.00 . In contrast, a similar analysis applied to the experienced-group scores showed the effect of prior probability to be highly significant, $F = 7.82$, $df = 1/269$, $p < .001$. Since the Prior Probability \times Ss interaction was not significant, $F = 1.27$, $df = 5/264$, $p > .25$, it was pooled with the error variability in order to gain df in the denominator of the F test for the prior-probability effect (Winer, 1962, p. 203). No other main or interaction effect was significant for the experienced group.

The results of the separate group analyses of the number of samples purchased are particularly interesting in that they provide an explanation for the discrepancy between the

observation of Green et al. (1964) that prior probability *does* influence action selection and Messick's (1964) finding that it does *not*. The essential difference may have been one of experience. The present findings, then, support the hypothesis that training is a prerequisite for sensitivity to prior probability.

Predictions of the More Frequent Datum

In order to assess the tendency to adopt the optimal strategy in the prediction of the next sample (maximization), a chi-square test was performed on the frequency of adoptions of the optimal strategy. From a total of 228 opportunities to adopt the optimal strategy within each group (i.e., 6 Ss \times 48 sequences), maximization occurred 225 times (or 78%) for experienced Ss as compared with 102 times (35%) for naive Ss. The difference was highly significant, $\chi^2 = 46.27$, $df = 1$, $p < .001$. However, this finding does not imply that the naive Ss' predictions were characterized by a probability-matching rule. A chi-square test of the naive Ss' predictions of the more likely sample against the expected frequencies indicated that even the naive Ss predicted the more likely sample more frequently than it actually occurred, $\chi^2 = 22.06$, $df = 1$, $p < .001$.

The overall finding of a tendency to maximize is not too surprising in view of the fact that payoff was used (see Luce & Suppes, 1965), and instructions specified that the sequence was randomly generated (McCracken, Osterhout, & Voss, 1962; Nies, 1962; Peterson & Ulehla, 1965). However, the fact that in the present study experienced Ss (trained under realistic circumstances) came much closer to a consistent maximization strategy than did naive ones suggests very strongly that real-life decision makers are less susceptible to maladaptive strategies (such as probability matching) than might be expected from studies conducted using college sophomores. Edwards (1961) has clearly demonstrated the importance of *specific* experience in probability-learning behavior; this study broadens the area of relevant experience to data evaluation and action selection.

Again, however, it should be pointed out that the present experienced Ss were members

of a team in an ongoing research project and, as such, may well have brought into the present situation characteristics other than sheer *decision-making experience* to differentiate them from naive Ss. Perhaps, for example, experienced Ss viewed the experiment as a temporary assignment and tended to accept the tedious task more willingly than did the naive Ss. Or, because of their more permanent employment (and higher rate of base pay), they may have been more highly motivated. In any case, it seems reasonable to argue that whatever may be confounded with *experience* in the present study, these same factors are likely to characterize real-life decision makers and their training. Therefore, perhaps the most important conclusion to be drawn from this work is not the importance of *experience* upon various facets of decision making, but the importance of *realistic training* in work purporting to represent general principles of decision behavior.

CONCLUSION

The data from the present study indicate that experience gained from training received under realistic circumstances is an important determinant of subsequent laboratory decision or choice behavior. First, the final subjective probabilities of the experienced group were consistently higher than those of the naive group for various levels of prior probability and proportion set. Second, the experienced group purchased fewer samples in each Prior Probability \times Proportion Set condition than did the naive group. Third, only the experienced group was affected by prior-probability values, strongly suggesting that experience acts as a moderator variable which increases the decision maker's sensitivity to such diagnostic cues. Finally, the experienced Ss tended to maximize (choose the more frequent event consistently) in a probability-learning situation to a greater extent than did the naive Ss.

All of these findings point to a general conclusion regarding the applicability of laboratory-based principles of decision making to real-life situations: unless *training* is realistic, the behavior of laboratory Ss may considerably underestimate human capabilities in a number of aspects of the decision process.

REFERENCES

- EDWARDS, W. D. Introduction to special issue on revision of opinions by men and man-machine systems. *IEEE Transactions on Human Factors in Electronics*, 1966, HFE-7, 1-6.
- EDWARDS, W. D. Probability learning in 1000 trials. *Journal of Experimental Psychology*, 1961, 62, 385-394.
- GREEN, P. E., HALBERT, M. H., & MINAS, J. S. An experiment in information buying. *Journal of Advertising Research*, 1964, 4, 17-23.
- IRWIN, F. W., & SMITH, W. A. S. Value, cost and information as determiners of a decision. *Journal of Experimental Psychology*, 1957, 54, 229-231.
- KAPLAN, R. J., & NEWMAN, J. R. Studies in probabilistic information processing. *IEEE Transactions on Human Factors in Electronics*, 1966, HFE-7, 49-63.
- KEMPTHORNE, O. The randomization theory of experimental inference. *Journal of the American Statistical Association*, 1955, 50, 946-967.
- LANZETTA, J. T., & KANAREFF, Z. T. Information cost, amount of payoff and level of aspiration as determinants of information seeking in decision making. *Behavioral Science*, 1962, 7, 459-473.
- LUCE, R. D., & SUPPES, P. Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Ed.), *Handbook of mathematical psychology*. Vol. 3. New York: Wiley, 1965.
- MCCRACKEN, J., OSTERHOUT, C., & VOSS, J. D. Effects of instructions in probability learning. *Journal of Experimental Psychology*, 1962, 64, 267-271.
- McHUGH, R. B. Comments on "Scales and statistics: Parametric and nonparametric." *Psychological Bulletin*, 1963, 60, 350-355.
- MESSICK, D. N. Sequential information seeking: Effects of the number of terminal acts in prior information. *Electronic Systems Division Technical Documentary Report*, 1964, No. 64-606.
- NIES, R. C. Effects of probable outcome information on two-choice learning. *Journal of Experimental Psychology*, 1962, 64, 430-433.
- PETERSON, C. R., & BEACH, L. R. Man as an intuitive statistician. *Psychological Bulletin*, 1967, 68, 29-46.
- PETERSON, C. R., & ULEHLA, L. J. Sequential patterns and maximizing. *Journal of Experimental Psychology*, 1965, 69, 1-4.
- SOUTHARD, J. F., SCHUM, D. A., & BRIGGS, G. E. An application of Bayes' theorem as a hypothesis-selection aid in a complex information-processing system. *USAF AMRL Technical Documentary Report*, August 1964, No. 64-51.
- SWETS, J. A., & GREEN, D. M. Sequential observations by human observers of signals and noise. In C. Cherry (Ed.), *Information theory*. London: Butterworth, 1961.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received March 13, 1968)

PUNITIVE SUPERVISION AND PRODUCTIVITY: AN EXPERIMENTAL ANALOG¹

DAVID R. SCHMITT²

Department of Sociology, University of Washington

This research concerns one unexplored aspect of the relationship between supervision and worker productivity—the manner in which the supervisor's activities are scheduled. A laboratory setting provided an analog to the supervisor's use of one type of consequence, punishment, to maximize the amount of time a worker spends in task activity while minimizing various unauthorized behaviors. The setting involved two concurrent operants reinforced with money where work on the higher paying one was penalized at various intervals. The effects of variations in the schedule of these intervals and the size of the penalties were explored. The results indicated that penalty magnitude significantly affected the allocation of work time when the penalties occurred at unequal intervals but not at equal ones. Under the unequal condition, the higher the penalties the less time spent on the punished task and the greater the time on the unpunished one. Low and moderate penalties, however, produced less work on the unpunished task than would be predicted on the basis of the possible losses through penalties.

As generally understood, supervision involves various activities which bear directly or indirectly on the job performance of the supervised individual: job planning, delegation of duties, communication of orders, and enforcement of work rules. The focus of a number of studies involving a variety of types of work groups has been the effects of the presence or absence of such activities or their combinations on worker productivity (Argyle, Gardner, & Coifi, 1957; Coch & French, 1948; Day & Hamblin, 1964; Gouldner, 1954; Katz, Maccoby, Gurin, & Floor, 1951; Katz, Maccoby, & Morse, 1950; Likert, 1961).

Supervision, however, is characterized by more than simply the presence or absence of various activities. The supervisor's choice of activities constitutes only one of the dimensions of what may be defined as his style of supervision. Of additional importance, although largely unexplored, may be the manner

in which these activities are scheduled. Two characteristics define the schedule of an activity—its *frequency* and its *regularity*. Thus any supervisory activity can occur at various frequencies and at intervals which may be either regular or irregular.

The potential effects of schedules would appear to be greater for some supervisory activities than for others. For activities such as job planning which usually occur infrequently and involve little interpersonal contact, the effects may be slight. However, for those which occur often and involve interaction between the supervisor and worker, the effects may be substantial. For example, a common function of supervision is to control the amount of work activity on an assigned job. In many settings supervisors "check up" on a subordinate to ensure that he is following his assignment. The importance of the frequency of such checkups has been suggested in research by Katz and his associates (Katz, Maccoby, Gurin, & Floor, 1951; Katz, Maccoby, & Morse, 1950) comparing the effects of close and general styles of supervision. In these studies supervisors of the less productive workers were found to be more likely to use close supervision involving frequent checkups and task instructions. In explanation, Kahn and Katz (1960) suggest that

¹This study was supported by the Cooperative Research Program of the Office of Education (Project No. S-319) and by the Graduate Research Committee of the University of Wisconsin. The author wishes to thank Lois Loddeke for her assistance in the research and L. Keith Miller and Robert Shotola for their suggestions and criticisms.

²Requests for reprints should be sent to the author, Department of Sociology, University of Washington, Seattle, Washington 98105.

most workers desire maximum autonomy and that supervision in a manner that does not permit it leads to lower morale and motivation. Other research, however, suggests that these effects may be limited to certain types of settings and production technologies (Argyle et al., 1957; Dubin, 1965).

Although unexplored, the regularity of the supervisory activities might have other important effects on work patterns. For example, in the use of supervisory checkups to ensure job performance, it might be predicted that regular checks will be less effective than irregular ones. With regular checkups the worker may learn when he needs to be present to coincide with the appearance of the supervisor, and thus may spend little additional time on the job. With irregular checkups, however, he may find such anticipation difficult or impossible, and thus must remain on the job for longer periods. Examples such as these suggest the potentially important effects of the schedule of an activity in supervisory situations and recommend its more systematic investigation in evaluating the effectiveness of various supervisory practices.

The general lack of research on schedule as an element of supervision style may have been dictated in part by the field research techniques that have typically been used in previous research on supervision. In general, field methods do not permit the measurement and control necessary to determine the effects which this aspect of supervision may have on productivity even though under some conditions it may determine the effectiveness of the supervisory activity.

The effects of the schedules of various consequences have been studied, however, in the experimental laboratory where sufficient measurement and control may be obtained. It may prove desirable, then, first to describe the effects of this variable experimentally, and then to determine the extent to which the results may be generalized to nonexperimental supervisory situations.

In the experimental study of task choice, a minimal task situation has been developed which permits the introduction of several conditions which appear to be functionally analogous to those in a nonexperimental situation involving the supervisor's use of check-

ups and sanctions to maximize the amount of time spent in work. The *S* in the experimental setting is confronted by two concurrent operants—spatially distinct tasks or responses simultaneously available to *S* (Catania, 1966; Ferster & Skinner, 1957). As with single operants, these tasks are simple, readily repeatable, and easily measured, for example, pressing a lever or button, pulling a knob. Different schedules of reinforcement or punishment are generally programmed for each of the operants. In such a multitask situation, various consequences may be manipulated to attempt to eliminate an individual's behavior on one of these operants while increasing it on a second. Such a condition appears to be functionally equivalent to the supervisor's use of various means to attempt to maximize the amount of time a worker spends in task activity while minimizing various unauthorized behaviors. While previous research in experimental psychology has explored some of the variables controlling concurrent behavior, unfortunately the combinations of conditions which might be generalized to a supervisory setting have not been studied.

This study attempts to demonstrate the manner in which the effects of one type of consequence, punishment, can be explored under conditions relevant to the study of the effectiveness of supervision. Punishment of various magnitudes was administered on two basic schedules for behavior on one of the two tasks. The study is the first in a series of laboratory experiments using variables analogous to various supervisory and task work conditions.

In its broadest sense punitive control includes a variety of punishing behaviors ranging from fines, threats, or physical abuse to more subtle acts such as criticism, ridicule, slights, snubs, or avoidance, and thus is manifest, at least to some degree, in almost all supervisory situations. This study focused on two variables relevant to the use of punishment in affecting the choice of activities: the magnitude of the punitive consequences and the schedule with which they are administered. Two types of schedules, fixed and variable interval, were explored. Studies of two task settings have not investigated the effects

of interval punishment on task choice. Rather, in previous research involving concurrent operants (Reynolds, 1963) or two choice risk-taking situations (Kogan and Wallach, 1967), punishment of one of the choices either has been continuous or has occurred for a particular proportion of the task responses. In general such studies suggest a tendency toward the elimination of the punished behavior as the negative consequences become high. The effects of fixed and variable interval schedules of punishment have been compared using a single operant (Azrin, 1956). These results indicate that variable interval schedules tend to produce more response suppression.

METHOD

Setting

The experimental setting in this study involved a choice of two activities each of which was reinforced. Both activities were button-pressing tasks located at opposite ends of a small work room. For each task, *S* was reinforced for pressing a large button mounted on an instrument panel. The reinforcer was money. A counter mounted on the panel indicated how much money *S* had earned. The tasks differed in the amount of money that could be earned on them. The number of presses required before a reinforcement count was registered was greater for one of the tasks. To standardize the rate at which different *Ss* could work on either task, a 3-sec. time-out occurred after each response. The number of responses for each cent earned on the higher paying task (Task B) was half that on the other (Task A). With four responses for each cent required on Task B, *Ss* could earn approximately \$2.80/hr; with eight responses required for each cent on Task A, *Ss* could earn \$1.40. Thus, of the two, Task B was the more attractive.

The effectiveness of the punitive consequences in changing task behavior was studied under conditions in which its interpretation would be relatively unambiguous. The consequences were evaluated regarding the degree to which they produced behavior on Task A, the less attractive task. Thus work on Task B, the more attractive task, was punished. In most nonexperimental settings the unauthorized activities which the supervisor punishes are probably not consistently more attractive than any other activity including the work itself, as in this study. Thus, if the consequences are effective in eliminating an activity which is considerably more attractive than any other situational alternative, they are likely to be at least as effective in other situations where the alternatives are of more equal attractiveness.

Work on Task B, the higher paying alternative, was periodically penalized by a loss of money. Only one of the two tasks was operable at a time. An

S-controlled switch on Task A determined which task could be used. The time at which work on Task B would be penalized was indicated by the sounding of a buzzer, regardless of which task *S* was operating. A penalty was administered only if *S* had Task B switched on when the buzzer sounded. A penalty count was added on a separate counter in the workroom; the amount of the penalty for that session was posted next to the counter. No consequences accompanied the buzzer if Task A was being operated. Since the changeover from work on Task B to Task A resulted in a several second delay while *S* crossed the room and turned on Task A, frequent switching to avoid penalties resulted in reduced reinforcement on either task. A clock on the wall was visible at all times. All events and measures were programmed and recorded by automated equipment in an adjacent room.

Procedure

The *Ss* were told only how to operate the tasks and that the sound of the buzzer would be followed by a loss of money if they were working on Task B. The *Ss* were college students who were told before volunteering that they would have an opportunity to make money on a laboratory task.

The effects of penalty magnitudes were explored under both fixed interval (FI) and variable interval (VI) schedules of supervision. Different *Ss* were used for each of the schedules. Within a schedule, however, *Ss* were exposed to several different penalty magnitudes. Changes in penalty were made only after *Ss* evidenced stability in task work under a given condition. Since this investigation focused on the extensive study of several *Ss* in each variation, a statistical analysis of the performance was judged not to be appropriate. Rather, similar patterns of response were sought in response to changes in the experimental conditions. The *Ss* worked in sessions of 1-4 hr. in length several times a week. Payment was made at the conclusion of the total hours of work.

RESULTS

Fixed Interval Supervision Schedules

Seven *Ss* worked over periods ranging from 4 to 14 hr. on several FI schedules in which the buzzer sounded after time periods of equal length throughout a work session. The different schedules included time intervals of 1, 3, 5, or 10 min. Penalties from \$.02 to \$2.00 were used. The *Ss* worked at least 1 hr. under each of the penalties.

The results indicated that none of the FI punishment schedules was effective in producing a substantial amount of activity on Task A. Figure 1 shows the percentage of time spent by five *Ss* on Task A working on

one of the schedules (FI 3 min.) under various penalty conditions. After less than 1 hr. of work under any of the schedules and penalty magnitudes, none of the Ss spent more than 30% of his time on Task A. With experience on a schedule, Ss avoided virtually all penalties by switching from Task A immediately after the buzzer and switching back again a few seconds before the next buzzer.

Variable Interval Schedule

Four Ss worked over periods ranging from 24 to 37 hr. on VI schedules in which the buzzer sounded after time periods of varying lengths. One schedule was used with an average of 4 min. for each interval. The intervals varied between 10 sec. and 8 min. Penalties varied from \$.01 to \$1.00 were used.

During the Ss' first 2 hr. of work on this schedule, no penalties were administered although the buzzer continued to sound at the various intervals. In the remaining hours for each S, one of two progressions of penalties was used. Two Ss were begun on high penalties which were progressively decreased when intersession stability was achieved. The other two Ss were begun on low penalties which were progressively increased. Several penalty

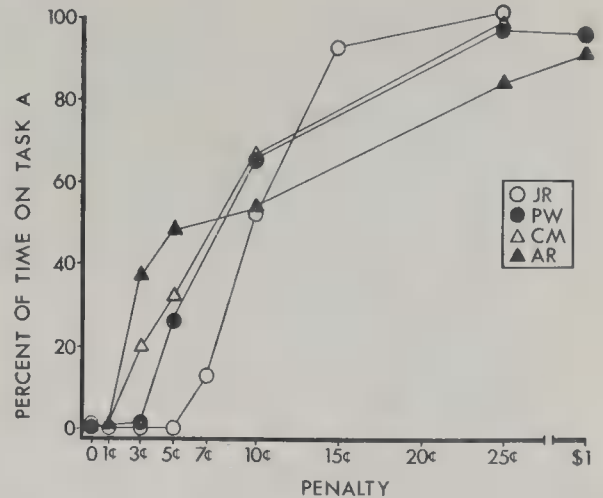


FIG. 2. Percent of time spent on Task A under various penalty conditions using a variable interval schedule (VI 4 min.). (The Ss JR and CM worked under progressively increasing penalties while AR and PW worked under progressively decreasing ones.)

magnitudes were repeated following intervening periods of work under other penalties to determine the replicability of their effects. The Ss worked at least 2 hr. under each penalty condition.

Figure 2 shows the proportion of time Ss spent on Task A under the various penalty magnitudes. The results indicate that VI punishment was effective in producing activity on Task A. For all Ss the proportion of time spent on Task A increased with increasing penalty size. Small penalties of less than \$.03 had a small effect on task behavior while moderate penalties from \$.05 to \$.15 considerably increased the time spent on Task A. High penalties of \$.25 or more generally resulted in time spent only on Task A after several hours of work. No pronounced effects appear to be caused by penalty sequence. Task performance under the various penalty conditions showed considerable stability and replicability particularly under the penalty extremes. For example, hourly differences in proportion of time on Task A under a given penalty averaged 9%.

DISCUSSION

The data clearly indicate the importance of different schedules in determining the effects of punishment on task choice. When penalties for work on one of the tasks were scheduled at equal intervals throughout a work period,

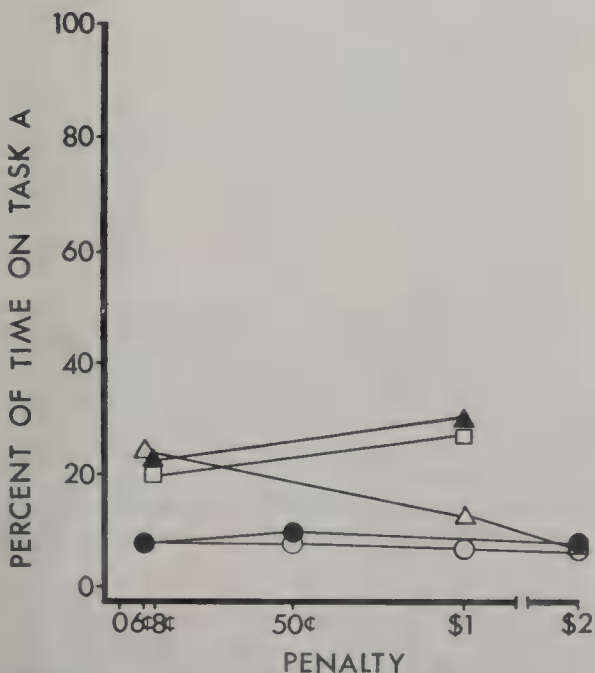


FIG. 1. Percent of time spent on Task A under various penalty conditions using a fixed interval schedule (FI 3 min.).

Ss learned quickly to avoid them. Thus regardless of their magnitude the penalties proved relatively ineffective in increasing activity on the unpunished task. In contrast, when the penalties were scheduled at unequal intervals, no S spent a large amount of time on the punished task without receiving a number of penalties. Under this condition, the larger the penalties the greater the time spent on the unpunished task.

Importantly, however, the effectiveness of VI penalties was not predictable as a direct function of their effect on total earnings. Low and moderate penalties produced more than the predicted amount of work on the unpunished task. The Ss tended to avoid losses often to the detriment of their total earnings. For example, Ss working on Task A earned approximately \$1.40/hr., on Task B \$2.80/hr. Thus with an average of 14–15 penalties randomly distributed per hour, Ss could maximize their earnings by working only on Task A with penalties greater than \$.10 and only on Task B with penalties less than \$.10. With \$.10 penalties, remaining on either task would result in approximately the same earnings. The results, however, indicate that with \$.03 penalties, only two Ss spent no time on Task A during these periods. The other Ss spent 20 and 38% of their time, respectively, on the lower paying task. With \$.05 penalties only one S spent no time on Task A with the other Ss spending 26, 32, and 49% of their time, respectively, on that task. For each of these penalty magnitudes the rank orders of the average amount earned by each S and the proportion of time spent on the higher paying task correspond exactly. With \$.10 penalties, the point at which either task could be selected with little difference in earnings, all Ss spent more than half of their time on Task A.

In conclusion, the inference drawn from these findings appears to be an important one for an analysis of the effectiveness of supervision. The results strongly recommend the consideration of not only the type of supervisory activity but also the schedules with which it is performed. As the case of punitive control illustrates, schedule type in conjunction with the magnitude of the punishment

may determine in large part the effectiveness of that activity.

A generalization of these results to non-experimental settings, however, should take note of the various limiting characteristics of this research. For example, punishment in the experimental setting was impersonal, specific to a given activity, and involved loss of money as the only aversive consequence. Much supervision in nonexperimental settings, however, is personal, associated with a number of poorly specified activities, and may involve a number of different consequences. In the experimental situation only two activities were available and money was used as the reinforcer for both, while in other settings workers often have many alternatives available which are reinforced in a variety of ways. In addition, workers on the job are often participants in formal or informal groups in which additional standards, pressures, or sanctions are imposed. To what extent such conditions alter the relationships found in this research in a "minimal" task situation will need to be determined.

REFERENCES

- ARGYLE, M., GARDNER, G., & COIFI, F. The measurement of supervisory methods. *Human Relations*, 1957, 10, 295–313.
- ARGYLE, M., GARDNER, G., & COIFI, F. Supervisory methods related to productivity, absenteeism and labor turnover. *Human Relations*, 1958, 11, 23–40.
- AZRIN, N. H. Effects of two intermittent schedules of immediate and nonimmediate punishment. *Journal of Psychology*, 1956, 42, 3–21.
- CATANIA, A. C. Concurrent operants. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application*. New York: Appleton-Century-Crofts, 1966.
- COCH, L., & FRENCH, J. R. P. Overcoming resistance to change. *Human Relations*, 1948, 1, 512–532.
- DAY, R. C., & HAMBLIN, R. L. Some effects of close and punitive styles of supervision. *American Journal of Sociology*, 1964, 69, 499–510.
- DUBIN, R. *Leadership and productivity*. San Francisco: Chandler, 1965.
- FERSTER, C. B., & SKINNER, B. L. *Schedules of reinforcement*. New York: Appleton-Century-Crofts, 1957.
- GOULDNER, A. W. *Patterns of industrial bureaucracy*. Glencoe, Ill.: Free Press, 1954.
- KAHN, R. L., & KATZ, D. Leadership practices in relation to productivity and morale. In D. Cartwright & A. Zander (Eds.), *Group dynamics*. (2nd ed.) Evanston, Ill.: Row, Peterson, 1960.
- KATZ, D., MACCOBY, N., GURIN, G., & FLOOR, L. G.

- Productivity, supervision and morale among railroad workers.* Ann Arbor: Survey Research Center, University of Michigan, 1951.
- KATZ, D., MACCOBY, N., & MORSE, N. C. *Productivity, supervision and morale in an office situation, Part 1.* Ann Arbor: Survey Research Center, University of Michigan, 1950.
- KOGAN, N., & WALLACH, M. A. Risk taking as a function of the situation, the person, and the group. In *New directions in psychology. III.* New York: Holt, Rinehart & Winston, 1967.
- LIKERT, R. *New patterns of management.* New York: McGraw-Hill, 1961.
- REYNOLDS, G. S. Potency of conditioned reinforcers based on food and on food and punishment. *Science*, 1963, 139, 833-839.

(Received March 14, 1968)

MARGINAL PRODUCTIVITY PROCEDURE FOR STAFF SELECTION¹

PURNELL H. BENSON²

Rutgers—The State University

A quantitative procedure for making staff selections is applied to the problem of hiring a research chemist. The procedure incorporates marginal productivity and marginal cost principles from economics, a numerical prediction score, and Bayesian principles of probability.

This article presents a procedure for selecting staff members in terms of an idea which is both old and new. The idea is an old one in that marginal theory in the purchase of the factors of production has existed for a century and a half, having originated with the British classical economists (Marshall, 1920). The idea appears new in that marginal productivity theory has become implicit in the staff-hiring decisions made by some personnel managers and executives whom the author has interrogated within recent years. This growing involvement gives the marginal productivity concept of hiring a face validity which justifies the effort to spell out explicitly what is assumed.

Brogden and Taylor (1950) maintain that whatever employee characteristics augment or decrease productive output can be accounted for in terms of dollars. Whoever at the supervisory level is best able to evaluate such factors in monetary value should do so. Haire (1959) points out that much more is involved in successful assignment than the dollar value of the employee's output, for example, job satisfaction, personality difficulties, grievances, job turnover, and the like. Dunnette (1963) rejects any single criterion of job success, and Wallace (1965) criticizes utility as a criterion.

That cost measures of the consequences of hiring continue to gain in popularity is reflected by studies such as those by Guttman and Raju (1965) and by Mahoney and England (1965). The present author recog-

nizes that productivity means much more than physical units of output from the individual employee. In this expanded sense, productivity is used in developing the selection procedure described here. The Bayesian concept of minimizing the weighted average risk (or of maximizing the weighted average of probable gains) is used (Birnbbaum & Maxwell, 1960; Cronbach & Gleser, 1965).

When the wage policy for employee classification is given to the personnel manager as a job specification, he endeavors to hire the best qualified applicants within the wage requirement. In a fixed-wage hiring decision, the marginal productivity theory of wages does not enter the problem of which individual to select.

The situation is different when the personnel manager is instructed to look for a staff member whose salary is open, depending upon qualifications, such as a salesman, researcher, engineer, manager, or executive. Then principles of marginal productivity guide the employment decision. The flexible-wage hiring decision is the type of employment decision reviewed here.

MARGINAL PRODUCTIVITY THEORY FOR SELECTING FACTORS OF PRODUCTION

Before proceeding with the mechanics of this decision, the principles of marginal productivity which involve staff selection may be briefly reviewed (Samuelson, 1953, 1964). A short statement of marginal analysis was given by Benson (1967).

Consider the two principal components of production: men and machines. If the available money for producing goods to meet company sales is \$1,000,000, this is divided between annual payroll for labor and annual

¹ A revision of material read by the author at the Operations Research Society of America in New York on May 31, 1967.

² Requests for reprints should be sent to the author, Graduate School of Business Administration, Rutgers—The State University, Newark, New Jersey 07102.

outlay for machines in that ratio which maximizes production. If there are too few men to operate machines, production falls. If there are too few machines, production likewise decreases.

The cost of production problem may be more easily visualized if the number of machines is considered fixed for the period of time under consideration. The question is asked: How large a labor force shall be employed to man the machines? Manifestly, as more men are hired, the output produced continues to rise. As men increasingly get in each other's way, output per man is reduced. The relationship between number of employees and amount produced is not a straight line, but rather a curve of diminishing increases in output as further men are added to the labor force already employed.

This concept of marginal productivity is readily shown by a graph of the value of the output produced. The curve in Figure 1 shows the increasing value of the total goods produced as more labor is purchased. As more goods are produced, their total value increases at a decreasing rate, owing to the increasing difficulty of making the goods with a limited number of machines in relation to the number of men.

Where the slope of the curve is 45° , the increase in the value of the goods is just equal to the cost of the last, or marginal, unit of labor bought. Below this point, the increment in value of goods sold is greater than the increment in the cost of labor. This indicates

that money is lost by not producing more. Above this point, the increment in the cost of labor is greater than the increment in the value of goods. This indicates that money is lost by spending more in producing goods than sales bring. The zero point of the graph is chosen where the cost of labor line meets the vertical axis.

THE DECISION TO ADD A STAFF MEMBER

Marginal productivity theory assumes that management is rational and knowledgeable. The theory sets forth an ideal for management to pursue in its quest to maximize financial return to those who provide capital for the enterprise. The provision of capital is also a factor of production controlled by marginal productivity theory.

To discover how the theory works, it is worth threading the line of administrative authority which eventually results in hiring a senior chemical analyst in the research laboratory.

The management of the company looks at the laboratory and decides how much research is needed for maximum sales at an acceptable margin of profit to stockholders. Management considers such things as number of new products needed, number of patents, number of improvements in manufacturing processes, and the effects of these things on augmenting sales. The head of the laboratory works within a budget. He is expected to produce as much research as possible within this budget. If he can produce research economically, management will conclude it can afford to undertake more research. The budget will be increased according to marginal productivity theory. If research becomes too costly in relation to what it brings back in sales, the budget may be reduced.

The research director builds up his staff with that mixture of abilities which will maximize the output of the laboratory within his allotted budget. He increases his staff only when he has authority to do so through a larger budget. He makes decisions from time to time in hiring replacements. These decisions are threefold. (a) How much money shall be spent to fill a job at the correct level of ability? (b) Which applicant best fits the job specifications? (c) What is the right salary to

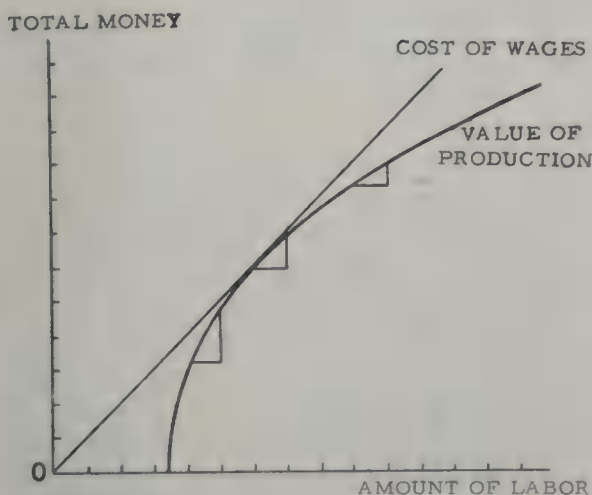


FIG. 1. Relationship between cost of wages and value of production.

pay the applicant in view of his particular qualifications? In this complex problem area the personnel department provides its research services to help the director.

SALARY LEVEL IN RELATION TO MARGINAL PRODUCTIVITY

Before a job can be filled, job specifications are prepared. Besides defining the kind of job to be filled, these specifications define a certain performance level. This level costs an expected annual salary which represents the market worth for a man of the required capability. The principal measure of the performance level is the salary cost in the employment market for procuring that level of performance.

For what performance level in a senior chemical analyst is it desirable to pay? If the man is merely to conduct routine analysis, it is not profitable to hire a brilliant and imaginative chemist at twice the salary needed for someone to do dependable analysis in the laboratory. It would also be unprofitable to hire an incompetent assistant incapable of complicated and accurate analysis. In either event, the output of the laboratory would suffer.

Whatever the units of output of the laboratory, the director judges how much it costs in different materials and skills for the last or marginal unit of output which is added by his administrative decisions. If he can hire an added man whose increment in output costs less than the same increment in output costs by using other factors, then he hires that man and uses less of other factors of production in his laboratory.

By hiring the optimum number of chemical analysts at performance levels which are most efficient for laboratory output, he is hiring each man at the point of expenditure where the man's contribution to laboratory output is just equal in value to the cost of hiring him. Here, the value of the increment in output is established by the prevailing costs of other factors of research needed to achieve such output. The increment in output is the change in output achieved by moving from one salary level to another. This must be estimated by the hiring officials or learned through research.

For maximum output within the available cost outlay each vacancy should be filled at that performance level where the output added by the employee compared with what is added by one of less capability at less salary just equals the wage difference paid the employee who fills the vacancy. The vacancy should not be filled at a lower level of performance, for the drop in output is greater than the saving in salary.

Nor should the vacancy be filled at a higher level of performance and salary. The gain in output is then less than the increment in salary cost. If the overhead expense of adding the man is simultaneously considered, the key to be evaluated is the difference in output between hiring Applicant A at one salary level plus overhead cost and hiring Applicant B at another salary level plus overhead cost.

It is convenient to graph the problem of performance level in terms of a net productivity curve. Figure 2 shows the curve of the value of the output added by the job of senior chemical analyst. The height of this curve is a function of increasing levels of performance, measured in salary terms. From this curve is subtracted the salary cost at each performance level. This yields the net productivity curve. The highest point of this curve is found where the slope of the productivity curve is 45° . The advantage in using the net productivity curve is that the hiring problem is more readily visualized as one of optimization. The highest point is achieved during job specification and applicant selection.

The ideas of management concerning changes in output resulting from different

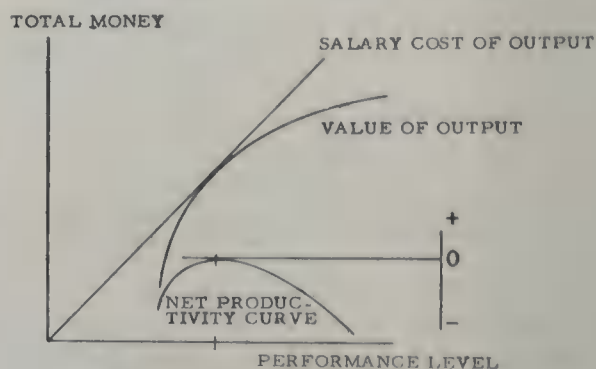


FIG. 2. Relationship between performance level and net productivity.

salary levels are vague at best. Management usually thinks in terms of what it would be just willing to pay as a bargain, if it could get someone whose performance level is too low or too high. The more rigorous approach is to think first of output resulting from different performance levels and then to subtract the salary which must be paid from the value of the output achieved. If necessary, the variable overhead costs for supervising a man of more or less capability should be included in calculating the net productivity curve.

Ideally, experimentation would be undertaken within a company to investigate the effect on productivity of assigning employees at various salary-performance levels. Short of an expensive experimental design, use may be made of historical data in which unplanned variations in assignment have occurred.

Usually it is necessary to rely upon estimates of net productivity made by managers. They have an idea of the optimum performance level at which to hire. The curve of diminishing returns is also required. To define this, supervisors should be asked, "If, as a bargain, you were able to hire an employee at the X-dollar level of performance, how much would you be willing to pay him?" From the judgments reported, the required curve can be drawn. A capable manager can answer in terms of the effect of salary level upon the output of the section, knowing what he should know about the marginal cost of securing changes in output through various administrative procedures.

TYPES OF INDIVIDUAL PERSONNEL DECISIONS

Several component decisions may now be differentiated.

1. *Job specification.* An aim of specification is to peg the salary-performance level at the point of highest marginal productivity to the company, after subtracting the salary paid. This is the point of highest profitability to the company. If the job is set too low, the company loses money through inefficient performance of the job. If the job is set too high, the company is paying for job quality it does not require. If the personnel manager is furnished the wrong job specification, he may find a man who is not good enough to

fill the job, or one who is costly because he is better quality material than the job actually requires.

2. *Individual selection.* The aim of the next step is to find the individual whose performance most closely matches the job description. If the job is filled with someone who does not meet the specified level properly, the effect of the error is the same as if the job were incorrectly specified before hiring. The job would then be filled with a man who is away from the optimum to one side or the other.

If the marginal productivity curve is relatively flat, it may make little difference whether the job is filled at optimum level. An error to either side would be minor in its effect. If the productivity curve, after subtracting the salary line, is quite peaked, a small error may have serious consequence. These considerations are illustrated by the curves in Figure 3.

3. *Paying the right salary.* If a manager pays too high a salary rate for his staff members, the effect is to lower the net productivity curve. The aim in hiring a single individual is to pay no more salary than is required to persuade him to accept employment. These details are illustrated in Figure 4. Figure 5 diagrams nine kinds of personnel choices. The three possibilities of paying more than, less than, or the right salary are combined with the three possibilities of hiring a man undergrade, overgrade, or at the right grade.

4. *Selecting the correct hiring opportunity.* The usual situation is one of several qualified applicants with somewhat different abilities and salary requirements. The aim is to select the applicant for whom the net productivity

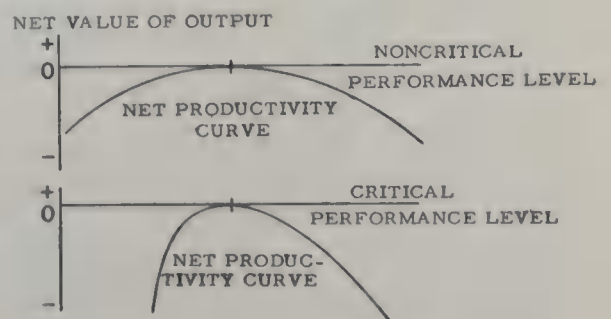


FIG. 3. Effects of noncritical and critical performance levels upon value of output.

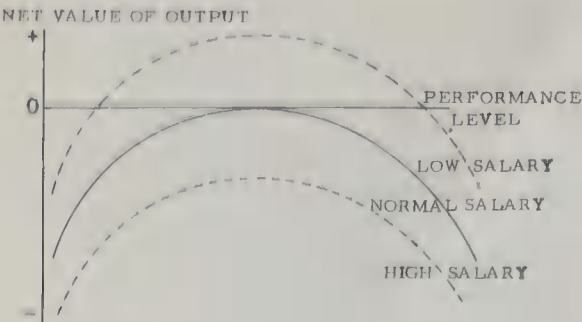


FIG. 4. Relationship between asking salary and net value of output.

is highest. A graph of the net productivity curve which marks the positions of applicants for the job is a useful aid in visualizing the hiring options. After the job curve has been drawn, the two pieces of information needed concerning the applicant are his performance level, measured in job-salary worth in the marketplace, and his salary requirement. His position on the graph relative to other applicants is then clear.

If an arbitrary limit has been imposed by management upon salary, then the aim is to select the best qualified applicant at the salary which can be paid. If an applicant can be hired for less than this, the choice is for the man whose net productivity is highest, within the salary restriction.

COST OF JOB REPLACEMENT

The cost of replacing an employee may be much greater than the drop-off from not filling specifications closely, or from paying him somewhat more than he is worth. When a man must be replaced, the cost to the company is illustrated in Figure 6. When the break-point is reached for replacing a man

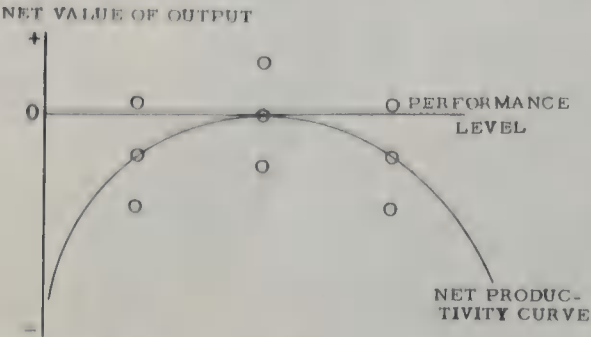


FIG. 5. Locations of net productivities of nine applicants with high, medium, and low performances and salaries.

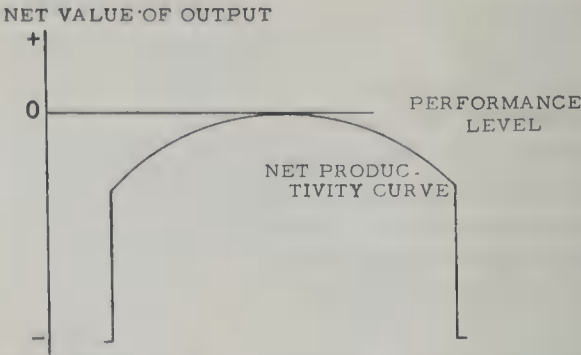


FIG. 6. Drop in net productivity curve of the job with job replacement.

who is under ability or over ability, the net productivity curve drops perpendicularly.

Past hiring experiences, both successful and unsuccessful, afford data for analyzing risks and costs in personnel decisions. Whether the employee makes good or not, and he usually does perform satisfactorily, follow-up research should ascertain how correctly the job was specified and whether the job specifications were properly met. Especially when a man is displaced, the personnel department should make inquiry to determine whether the job was correctly described and filled, and whether at the correct level of compensation.

ROLE OF THE PREDICTION SCORE IN MARGINAL PRODUCTIVITY HIRING

A numerical score for predicting job success combines numerical weights established through multiple regression analysis of variables associated with job success in past cases.

Applied to the marginal productivity procedure for hiring, the prediction score should estimate the market worth of the individual in terms of salary for filling the specified job. Then the productivity graph, as in Figure 5, can be marked on the horizontal axis to show how near the optimum the individual falls on the net productivity curve. At the same time, the numerical score indicates the proper salary to be paid to the prospective employee. Depending upon whether he accepts a somewhat higher or lower salary, his net productivity is below or above the curve.

The dependent variable in the multiple-regression analysis leading to the prediction score is salary paid in relation to performance. In a general sense, the salary received by

present employees can be used as the dependent variable. Alternatively, estimates by associates of the salary which each individual is worth may be used.

The prediction score contains errors from sampling traits of the individual, errors from sampling populations of employees, errors in ratings, and errors of estimation resulting from the incomplete correlation between predictor variables and the predicted numerical score. The prediction score does not give a knife-edge estimate of performance in market salary terms, but a statistical distribution of probable levels of performance. Corresponding to the distribution of levels is a distribution of expected net productivities from hiring the individual. The summation of these separate productivities times the chances of their occurrence gives the expected productivity return from hiring the individual. This applies Bayesian decision theory.

If the numerical functions involved are continuous and adequately described, exact integration can be performed. Otherwise, approximation methods must suffice. A procedure which combines knowledge of the standard error of estimate of the numerical score and knowledge of the net productivity curve is illustrated in Figure 7. The expected productivity from hiring the applicant is equal to the average of all of the productivities within the error distribution. An approximate procedure marks off, by vertical lines, the normal error distribution into 10 or 20 equal areas. The mean of the net productivities for the midpoints in each of these intervals estimates the expected productivity from hiring the applicant.

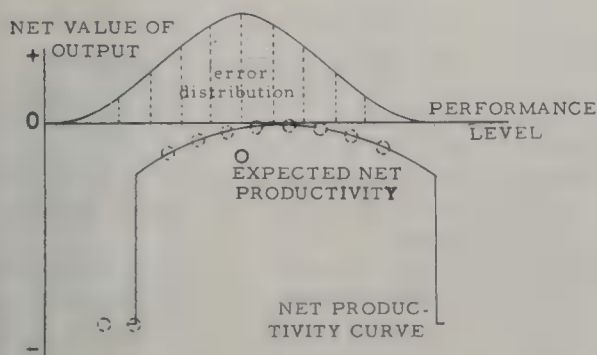


FIG. 7. Calculation of expected net productivity when the distribution of errors in estimating performance is known.

From the foregoing, it is apparent that job success is not a categorical question, but rather one of relative financial return to the company. If the individual's own future can be weighed in monetary terms, this also can be included in the comparison of alternative job possibilities.

If little is known about the applicant, the standard error of his prediction is large. The error distribution then encompasses both the risks of hiring someone inadequate for the job or too costly for the job. The negative productivity values from the vertical drop-off on either side of the net productivity curve then result in much smaller expected value to the company from hiring the applicant.

After the experience has shown how much reduction in the standard error of the prediction score can be gained from research outlay, then it can be decided if it is financially desirable to strive for greater precision in the prediction score.

SUMMARY OF STEPS IN THE MARGINAL PRODUCTIVITY HIRING PROCEDURE

1. The job vacancy is fully specified regarding both the abilities and qualifications of the person required to fill it and the estimated salary in the marketplace of a person capable of filling it at the optimum salary level. (If the job vacancy is a position which a trainee is expected to fill in the future, the analysis is built around the future assignment. In this case the procedure is a logical extension of that given for filling an immediate position.)

2. A graph is constructed of the net productivity curve. This curve is based upon experimental or historical data, if available. Otherwise, the curve is located by asking the department head and his associates for estimates of the net worth to the department of filling the job vacancy at various salary levels.

3. The abilities and qualifications of each applicant are described in detail from resumes, references, transcripts, and interviews.

4. An individual estimate is made of the market value of each applicant in salary terms. How much are his abilities worth when measured in terms of what the company expects to pay? This estimate can be made from multiple-regression analysis of salaries

and qualifications of present personnel and their job specifications, or from intuitive judgment of the personnel officer, or from both of these sources of information. (If the estimate is recognizably approximate, then the distribution of errors of estimate must be considered in making an average estimate of the possible results.)

5. The minimum salary which the applicant will accept is ascertained.

6. The individual's position on the net productivity graph is plotted by locating his market salary worth as the horizontal coordinate. The vertical coordinate is given by subtracting the surplus of his salary requirement over his salary worth from the net productivity curve. The height of this position gives the net productivity of the applicant.

7. Of the various applicants, that one is hired whose coordinate position shows the highest net productivity. If a salary limitation has been imposed, then that applicant is hired who falls within the salary limitation and whose net productivity is highest.

Example. A job vacancy exists for a senior chemical analyst. Supervisors report that the job would ideally be filled by a \$12,500-a-year man. If filled by a \$10,000-a-year man instead of a \$12,500-a-year man, the department would be willing to pay only \$8,000 for him in terms of anticipated department output priced at prevailing costs for such output. If the job is filled by a \$15,000-a-year man, the department would feel it worth paying only \$13,500 for what he could do in the job opening.

These details fix the net productivity curve as down \$2,000 at the \$10,000 horizontal coordinate, and down \$1,500 at the \$15,000 horizontal coordinate. Taking the origin at the peak of the curve, \$12,500, the formula fitted to these three points is $Y = -.28X^2 + .016X^3$, where Y is the net productivity in thousands, and X is the salary worth in thousands. The resulting curve is drawn in Figure 8.

Individual A has qualifications which in the opinion of hiring officers are worth \$11,000 in the marketplace for one who can fill the job opening. His asking salary is \$11,500. The difference between these two figures is \$500. His market worth for qualifications

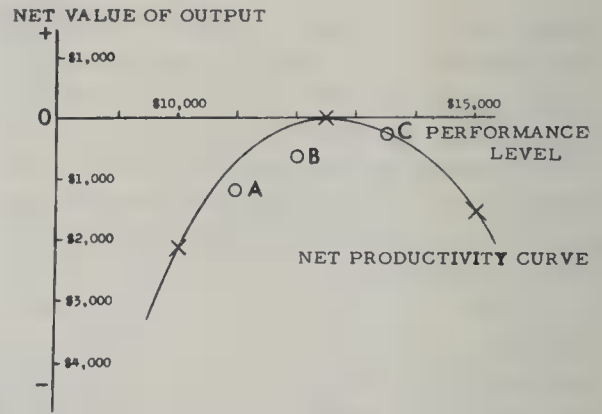


FIG. 8. Calculation of net productivities of three applicants.

useful in the job opening locate him along the horizontal at \$11,000. His salary requirement is \$500 in excess of this, which locates him \$500 under the net productivity curve and defines his net productivity as $(-\$684 - \$500) = -\$1,184$. Similarly, the net productivities of Applicants B and C are found from their data to be $-\$572$ and $-\$264$. Applicant C, having the highest net productivity, is hired.

It will be noted that the net productivities are measured in relation to what would be the maximum departmental output if the job vacancy were exactly filled by one whose salary requirement is equal to his salary worth. This provides the most convenient origin for carrying through the computation. The profitability of the department to the company is larger than this. If the profitability of the department were not believed to be positive, its budget would be altered by management. The curves are drawn with reference to net value of output from hiring the applicant.

REFERENCES

- BENSON, P. H. Multiple-regression analysis of a paired-choice division-of-time inventory in relation to grade-point average. *Journal of Applied Psychology*, 1967, 51, 82-88.
- BIRNBAUM, A., & MAXWELL, A. E. Classification procedures based on Bayes' formula. *Applied Statistics*, 1960, 9, 152-169.
- BROGDEN, H. E., & TAYLOR, E. K. The dollar criterion—applying the cost accounting concept to criterion construction. *Personnel Psychology*, 1950, 3, 133-154.
- CRONBACH, L. J., & GLESER, G. C. *Psychological tests and personnel decisions*. Urbana: University of Illinois Press, 1965.

- DUNNETTE, M. D. A note on the criterion. *Journal of Applied Psychology*, 1963, **47**, 251-254.
- GUTTMAN, I., & RAJU, N. S. A minimum loss function as determiner of optimal cutting scores. *Personnel Psychology*, 1965, **18**, 179-185.
- HAIRE, M. Problems relevant to business and industry. *Psychological Bulletin*, 1959, **56**, 169-194.
- MAHONEY, T. A., & ENGLAND, G. W. Efficiency and accuracy of employee selection decision rules. *Personnel Psychology*, 1965, **18**, 361-377.
- MARSHALL, A. *Principles of economics*. New York: Macmillan, 1920.
- SAMUELSON, P. A. *Foundations of economic analysis*. Cambridge: Harvard University Press, 1953.
- SAMUELSON, P. A. *Economics: An introductory analysis*. New York: McGraw-Hill, 1964.
- WALLACE, S. R. Criteria for what? *American Psychologist*, 1965, **20**, 411-417.

(Received March 15, 1968)

ACHIEVEMENT IN SCIENCE AND INTELLECTUAL STYLE

A. J. CROPLEY¹

AND

T. W. FIELD

*University of Saskatchewan,
Regina Campus, Canada*

University of New England, Australia

Some clever high school students are academically successful, but others with equally high IQs are not. Thus, it is possible that achievement involves cognitive variables other than simply level of intellect (IQ). Science achievement was shown to be related to several such variables including abstract thinking, originality, and category width. These findings were interpreted as supporting the importance of "intellectual style" in achievement. Large sex differences were also obtained. Better understanding of the role of style in classroom performance would be particularly useful in view of the current desire to identify talent early and to foster its realization.

Not all people with high IQs make outstanding contributions to their particular areas of specialization. Terman's geniuses, for example, although much more successful as a group than the average, were not universally characterized by exceptional intellectual contributions in the various fields they entered (Terman & Oden, 1959). Such substantial, but far from perfect, relationships between level of ability and performance are found in science too. Despite increased interest in fostering scientific achievement in recent years, there is still no clear understanding of how potentially successful science students differ from other able students who either do not enter science, or who enter the field but do not succeed. It is now apparent that mere level of ability (usually expressed in the form of an IQ) is not the key factor which distinguishes the science specialist from the non-scientist, nor the successful science candidate from the unsuccessful. Gibson and Light (1967), for example, were unable to distinguish among Cambridge University scientists, or to distinguish them unequivocally from nonscientists on the basis of IQ. Although level of ability is clearly relevant, perhaps in the form of a "threshold" (McClelland, 1958, pp. 12-13) below which effective scientific thinking is not possible, not all successful scientists have unusually high IQs, nor are all would-be scientists with high IQs successful.

Increasingly nowadays the notion of intellectual "style" or "bias" (Hudson, 1966), as

against mere level, is being emphasized as an important variable in the study of intellectual processes. The concept of style refers to the existence of certain stable idiosyncratic differences among people in the way in which they go about taking in, processing, and utilizing information obtained from their environments (Schroder, Driver, & Streufert, 1967). Two people of equal capacity may differ markedly in the characteristic ways in which they deploy their intellectual resources in coming to grips with information and in the kinds of information which they prefer to handle. The concept of style is, in fact, now well established in the literature of cognitive psychology (Ausubel & Ausubel, 1966; Gardner, Holzman, Klein, Linton, & Spence, 1959; Witkin, 1964), but less frequently utilized in the applied field.

Furthermore, there is some empirical evidence that preference for science as against the arts (Hudson, 1963a, 1963b) and outstanding achievement in science (Cropley, 1967b) are related to intellectual style, while theory and popular stereotype both suggest that there is a distinct cluster of style variables characterizing scientists (Barron, 1965, pp. 85-86). Consequently, the present study was concerned with the question of whether high science achievers at senior high school level did differ markedly from low achievers in terms of style, as against level, of intellect. The relationship of science achievement to IQ is well known (Cline, Richards, & Needham, 1963). The present study was aimed at answering the question of whether differ-

¹Requests for reprints should be sent to A. J. Cropley, Division of Social Sciences, University of Saskatchewan, Regina, Saskatchewan.

ences of a more qualitative kind exist between achievers and nonachievers.

METHOD

Subjects

A battery of tests was administered to all fifth and sixth form students in two high schools in large country towns in New South Wales, Australia. A total of 178 science students ranging in age from 15 yr. 8 mo. to 18 yr. 7 mo. completed all tests. This group included 104 boys with a mean age of 17.09 yr. ($SD = .64$ yr.), and 74 girls for whom the mean age was 16.89 yr. ($SD = .61$ yr.).

Tests

The battery of tests included a standardized test of science achievement specifically designed for use with Australian students at this level, published by the Australian Council for Educational Research (ACER). A measure of level of ability (IQ) was also obtained, using the AL-AQ test of intelligence. This test, again published by ACER, is intended for use with students at senior high school level and above, and so avoids some of the variance restriction inherent in a highly selected sample like the present one. Finally, four tests involving what is here called intellectual style were administered. These four tests included tests of originality (Torrance, 1962), flexibility (Torrance, 1962), category width (Pettigrew, 1958), and a test of the abstractness of intellectual functioning based on the developmental psychology of Piaget (Tisher, 1962). This test involved showing the students relatively common situations, for example, two partly filled containers of water linked by a tube. The students were asked both to predict the results of certain perturbations introduced into the situations, for example, raising one container above the other, and, given a result, to say what perturbation would have been necessary to yield that

result. Scoring was based, not on the rightness or wrongness of answers, but on the extent to which answers were of a concrete or formal kind. Thus, the Tisher test is a version of the Piaget interview technique (e.g., Inhelder & Piaget, 1958). Tisher (1962) reported high agreement between ratings obtained actually using the standard interview technique and ratings yielded by his group procedures.

Procedure

Standard tests were scored according to the published specifications. The originality score was obtained by scoring the Tin Can Uses test according to the differential weighting procedure described by Torrance (1962) and using the weights suggested by Cropley (1967a, pp. 109-110). The flexibility score involved rescoring the same test, this time allotting a point for each clear switch of topic in a given student's responses to a particular item (Torrance, 1962). Finally, the scores for abstractness of thinking were obtained from ratings of students as either predominantly concrete, early formal, or late formal, in their style of responding to the Tisher test. Students were similarly trichotomized on the other three style variables too, by dividing them into high, middle, and low scorers. As nearly as tied scores permitted, each of these groups contained an exact third of the total group. Subsequently, the three groups were further subdivided according to sex. A two-way analysis of covariance was then carried out on the science achievement scores of the various style groups, with IQ as covariate in each case, following the model proposed by Winer (1962, pp. 590-600).

RESULTS

Product-moment intercorrelations, means, and standard deviations for achievement, IQ, and style variables are shown in Table 1. Where data are available, reliabilities are

TABLE 1
INTERCORRELATIONS, MEANS, STANDARD DEVIATIONS, AND RELIABILITIES
FOR ALL VARIABLES

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|----------------------|------|---------|---------|------|-------|------|
| 1. Abstract thinking | — | -.08 | -.01 | .02 | .37 | .39 |
| 2. Originality | .17 | — | .74 | .21 | .07 | -.05 |
| 3. Flexibility | .16 | .61 | — | .23 | .22 | .01 |
| 4. Category width | .03 | -.02 | .06 | — | .11 | -.01 |
| 5. IQ | .49 | .09 | .04 | -.03 | — | .64 |
| 6. Achievement | .49 | .25 | .18 | -.21 | .67 | — |
| Males | | | | | | |
| <i>M</i> | 2.29 | 48.9 | 16.6 | 66.3 | 117.2 | 47.5 |
| <i>SD</i> | 0.78 | 19.5 | 3.28 | 14.7 | 10.3 | 8.34 |
| Females | | | | | | |
| <i>M</i> | 1.77 | 37.5 | 14.8 | 54.3 | 117.4 | 44.3 |
| <i>SD</i> | 0.75 | 16.6 | 3.32 | 16.1 | 10.2 | 8.35 |
| Reliability | — | .64-.71 | .60-.62 | .72 | .88 | — |

Note.—With $df = 72$, critical value of a correlation coefficient ($p < .05$) is .225; with $df = 102$, the value is .191.

TABLE 2
MEANS, STANDARD DEVIATIONS, AND FS FOR ACHIEVEMENT

| Category | High | | | Middle | | | Low | | | Fs ^a | |
|----------------|----------|-----------|----------|----------|-----------|----------|----------|-----------|----------|-----------------|-------------|
| | <i>M</i> | <i>SD</i> | <i>N</i> | <i>M</i> | <i>SD</i> | <i>N</i> | <i>M</i> | <i>SD</i> | <i>N</i> | Main effect | Interaction |
| Operations | | | | | | | | | | | |
| Males | 52.2 | 7.49 | 33 | 45.7 | 8.03 | 61 | 42.6 | 5.49 | 10 | 3.90* | 0.13 |
| Females | 52.2 | 8.37 | 11 | 44.6 | 7.16 | 43 | 39.5 | 7.15 | 20 | | |
| Originality | | | | | | | | | | | |
| Males | 47.8 | 9.63 | 47 | 48.4 | 8.47 | 31 | 47.8 | 7.83 | 26 | 1.09 | 3.10* |
| Females | 46.6 | 8.44 | 13 | 44.9 | 6.82 | 28 | 42.7 | 8.53 | 33 | | |
| Flexibility | | | | | | | | | | | |
| Males | 48.0 | 8.92 | 39 | 46.5 | 7.91 | 43 | 48.3 | 7.88 | 22 | 0.66 | 2.65 |
| Females | 45.3 | 8.23 | 14 | 47.2 | 7.98 | 23 | 42.4 | 8.09 | 37 | | |
| Category width | | | | | | | | | | | |
| Males | 46.8 | 7.88 | 49 | 48.6 | 9.20 | 31 | 47.4 | 7.90 | 24 | 5.91** | 1.43 |
| Females | 40.7 | 5.65 | 11 | 43.9 | 9.80 | 28 | 46.0 | 7.43 | 35 | | |

^a For all Fs, *df* = 2/171.
* *p* < .05.
** *p* < .01.

also shown. The figures for originality, flexibility, and category width are test-retest reliabilities, while for IQ they are Spearman-Brown split-half coefficients. The two figures for originality and flexibility indicate the range of reliabilities cited in a number of reliability studies reported in the test manual. Correlation data above the diagonal in Table 1 are those for boys, data below the diagonal are those for girls.

Results of the analyses of covariance of sex, style, and achievement data are shown in Table 2. This table also shows achievement means for the various style groups and numbers in those groups. Relationships among style, achievement, and sex were significant for three style variables out of four.

The particular procedure adopted in forming the style groups made it possible to test the significance of differences between males and females in their distribution among the high, middle, and low categories for each style variable. Relationships were significant in all cases, the χ^2 value for operations being 12.7 (*df* = 2, *p* < .01), for originality, 15.8 (*df* = 2, *p* < .01), for flexibility, 17.2 (*df* = 2, *p* < .01), and for category width, 22.0 (*df* = 2, *p* < .01).

DISCUSSION

These results indicate that, at least as far as the present students were concerned, cognitive variables which may loosely be charac-

terized as describing style rather than level of intellectual performance accounted for significant portions of the variance of science achievement. The data thus support the notion that one of the reasons why only some able people become successful scientists may be because there is a particular kind of cognitive organization appropriate to science achievement after level (IQ) effects have been removed. They also provide some suggestions of what such organizations involve. The most successful science students in the present study were characterized by highly abstract and original thinking and by their characteristic ways of relating apparently discrepant data. These findings are consistent with others based on scores of unusually successful undergraduate scientists in an Australian university (Cropley, 1967b). In a longitudinal study covering the 4 yr. required for an undergraduate honors course in science, it was shown that men graduating with honors came almost exclusively from among those who had been rated highly divergent in their style of thinking on entry to the university 4 yr. previously.

Of course the present data do not show that the relationships between achievement and what is here called intellectual style are peculiar to science achievement. It is possible that the same relationship holds for achievement in any area. However, Hudson (1963a, 1963b, 1966) has demonstrated the existence

of marked differences between arts and science specialists in terms of intellectual bias. His findings, which are well established as far as English grammar school boys are concerned, show that differences exist between science achievers and nonscience specialists in areas similar to those involved in the present study. Thus, although the possibility of a general and simple relationship between style of intellect and achievement cannot be discounted on the basis of the data reported here, other studies suggest that this may not be the case.

Sex differences on the cognitive variables are interesting in the light of a report of Broverman, Klaiber, Koboyashi, and Vogel (1968) that differences in cognitive performance between males and females may well reflect physiological rather than sociological differences between the sexes, centering on the greater capacity of males for inhibition of ongoing overlearned behaviors in favor of more complex, original responses. Such differences were not revealed by the IQ test, on which the boys' mean of 117.2 ($SD = 10.3$) was almost identical to the girls' mean IQ of 117.4 ($SD = 10.2$). In the main, too, the sex differences were in the direction of superior achievement by males.

Since the tests were all administered at the same time, it is not clear whether the existence of cognitive organizations appropriate to high science achievement led certain individuals to enter science in the first place, or whether effective science training resulted in the increased scores of high achievers on the style variables. Possibly, nonintellective factors, for example, temperament, desire for status, parental pressure, etc., induce students to enter science, success going to those who acquire most successfully the necessary style of functioning as a result of their scientific training. The present data do not indicate which is the case. Nonetheless, the longitudinal study of Australian undergraduates already mentioned showed that potential honors graduates were already identifiable in terms of intellectual bias 4 yr. before graduation, at a time when neither achievement nor IQ scores differed from those of ultimately less successful students. In any case, the suggestion that intellective variables other than IQ are significantly connected with academic success is

particularly relevant to current concern with early recognition of children with unusual talent.

REFERENCES

- AUSUBEL, D. P., & AUSUBEL, P. Cognitive development in adolescence. *Review of Educational Research*, 1966, **36**, 403-413.
- BARRON, F. *The psychology of creativity. New directions in psychology: II*. New York: Holt, Rinehart & Winston, 1965.
- BROVERMAN, D. M., KLAIBER, E. L., KOBAYASHI, Y., & VOGEL, W. Roles of activation and inhibition in sex differences in cognitive abilities. *Psychological Review*, 1968, **75**, 23-50.
- CLINE, V. C., RICHARDS, J. M., & NEEDHAM, W. E. Creativity tests and achievement in high school science. *Journal of Applied Psychology*, 1963, **47**, 184-189.
- CROPLEY, A. J. *Creativity*. London: Longmans, 1967.
- (a)
- CROPLEY, A. J. Divergent thinking and science specialists. *Nature*, 1967, **215**, 671-672. (b)
- GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, H. B., & SPENCE, D. P. Cognitive controls. *Psychological Issues*, 1959, **1**, 1-186.
- GIBSON, J., & LIGHT, P. Intelligence among university scientists. *Nature*, 1967, **213**, 441-443.
- HUDSON, L. Personality and scientific aptitude. *Nature*, 1963, **198**, 913. (a)
- HUDSON, L. The relation of psychological test scores to academic bias. *British Journal of Educational Psychology*, 1963, **33**, 120-131. (b)
- HUDSON, L. *Contrary imaginations*. London: Methuen, 1966.
- INHELDER, B., & PIAGET, J. *The growth of logical thinking*. New York: Basic Books, 1958.
- MCCLELLAND, D. C. Issues in the identification of talent. In D. C. McClelland, A. L. Baldwin, U. Bronfenbrenner, & F. L. Strodbeck (Eds.), *Talent and society*. Princeton: Van Nostrand, 1958.
- PETTIGREW, T. F. The measurement and correlates of category width as a cognitive variable. *Journal of Personality*, 1958, **26**, 532-544.
- SCHRODER, H. M., DRIVER, M. J., & STREUFERT, S. *Human information processing*. New York: Holt, Rinehart & Winston, 1967.
- TERMAN, L. M., & ODEN, M. *Genetic studies of genius*. Vol. 5. Stanford: Stanford University Press, 1959.
- TISHER, R. P. The development of some science concepts: A replication of Piaget's studies. Unpublished bachelor's (honors) thesis, University of New England, Australia, 1962.
- TORRANCE, E. P. *Guiding creative talent*. Englewood Cliffs, N. J.: Prentice-Hall, 1962.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- WITKIN, H. A. Origins of cognitive style. In C. Scheerer (Ed.), *Cognition: Theory, research, promise*. New York: Harper & Row, 1964.

(Received March 18, 1968)

PATTERN ANALYSIS OF BIOGRAPHICAL PREDICTORS OF SUCCESS AS AN INSURANCE SALESMAN

ROBERT TANOFSKY, R. RONALD SHEPPS,¹ AND PAUL J. O'NEILL²

Sales Personnel Research, Metropolitan Life Insurance Company, New York City

The investigators attempted to examine, with the aid of a computer program, all possible combinations of biographical items associated with different levels of sales production. The Ss were 1,525 life insurance salesmen. A combination of high prior income and more than two dependents at the time of application was found descriptive of high producers. Low producers were characterized by low earnings prior to their appointment. Age, education, marital status, and sales experience were of negligible importance.

The present investigation is part of a long-term effort to clarify the relationship between biographical characteristics of life insurance salesmen and their actual sales performance. Studies by Ferguson (1960) and Shepps, Tanofsky, and Mead (1967) have shown economic history to be predictive of sales production during the first 1-3 yr. on the job. The magnitude of this relationship is suggested by the correlation of .32 observed in the latter study between salary and first year sales.

In the present study it was decided to examine jointly salary and five other biographical predictors. Education, number of dependents, marital status, age, and previous sales experience have all been studied individually by the Life Insurance Agency Management Association (1954). They all are reported to be predictive of success as a life insurance agent, although in statistically complex ways. It was decided to examine these predictors by means of pattern analysis, a technique for the inspection of the combined pattern of scores contained within a set of predictors.

Meehl (1950) gave an early demonstration of the manner in which the pattern formed by two dichotomized predictors could show very high validity despite the zero validity of the

predictors taken singly. Using two trichotomized MMPI scales in combination, Lykken and Rose (1963) were able to make more accurate predictions than was possible with the conventional multiple-regression formula using all 11 predictors. A review of the clinical use of the technique of pattern analysis is now available (Sines, 1966).

Of more immediate interest is Sorenson's (1964) investigation. In one of the rare uses of the configural approach within industry, Sorenson found he could make more accurate predictions of sales effectiveness using patterns of personal characteristics taken four at a time than was possible for any weighted composite of biographical items taken one at a time. In other words, Sorenson's data suggest that combinations of personal history items can predict a sales criterion more successfully than can the classical linear-regression technique. Dunnette's (1966) theoretical model of the ideal personnel selection procedure also calls for the use of a combination of predictors to select those applicants most likely to succeed.

Since the foregoing literature gave no indication of the patterns to be expected from the six variables under investigation, the more general hypothesis was tested that specific patterns would be found which would be related to specific levels of production.

METHOD

Procedure

A computer program was used to examine rapidly all possible combinations of the 26 intervals associated with the six predictive variables (Table 1). Sonquist and Morgan's (1964) program was in fact

¹ Requests for reprints should be sent to R. R. Shepps, Marketing and Field Management, Metropolitan Life Insurance Company, 1 Madison Avenue, New York, New York 10010.

² The authors wish to acknowledge A. Trokenheim's contribution to the early portion of the present research. They also wish to express appreciation to the Service Bureau Corporation, New York City, for their help in securing the pattern analysis.

TABLE 1
CATEGORIES OF THE PREDICTOR VARIABLES

| Variable | Category | | | | | |
|---------------------|-----------------------|--------------------|-----------------------|-------------------|-------------------|-------------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Age | Under 25 yr. | 25-29 yr. | 30-39 yr. | Over 39 yr. | | |
| Prior weekly salary | Less than \$90 weekly | \$90-109 weekly | \$110-129 weekly | Over \$129 weekly | | |
| Sales experience | Tangible only | Intangible only | Both kinds | None | | |
| Marital status | Single | Married | Other | | | |
| Dependents | 0 | 1 | 2 | 3 | 4 | More than 4 |
| Education | Grade school only | Attend high school | Graduated high school | Attend college | Graduated college | |

written especially for the many biographical and demographic variables whose intervals consist of simple categories or classifications, such as "Married," "Single," and "Other" for the marital status variable. The program did not make any assumptions of linearity or additivity. Combinations of intervals were reported in the order in which they accounted for sales production variance. In effect, the program divided the total group of Ss into mutually exclusive subgroups of increased homogeneity with respect to the criterion. Such partition required at least 25 Ss in each group. The minimum sum of squares of the group must be at least 1.5% of the total sum of squares. In order for a group to be formed, a minimum reduction of .5% in the error sum of squares also had to occur. The program automatically stopped when it was unable to use the remaining intervals to divide further any segment of the sample.

It must be stressed that the computer simply used at every point the most important of the remaining categories, that is, the one which would bring about the greatest reduction in the remaining error sum of squares. Statistical significance was not taken into consideration. Sonquist and Morgan (1964) observe that "It seems unreasonable to apply ordinary statistical tests at each split; that is, to insist that the split be a statistically significant difference between the two means. It is the best of a large number of splits at each stage [p. 114]." The present investigators nevertheless applied statistical tests of significance to the outcome of the pattern analysis to determine whether the groups that emerged were reliably different from each other on the criterion measure, and also to determine whether the program had succeeded in accounting for a significant amount of the total variance.

Subjects

The Ss for this investigation were 1,525 male agents appointed during the year 1963 by Metropolitan Life Insurance Company who survived at least four quarters after the time of appointment. The biographical information on each salesman was

obtained from the application blank he filled out at the time he applied to the Company.

RESULTS

The results of the computer program for pattern analysis are reported in Figure 1. Prior income and number of dependents accounted for virtually all the explicable production variance. Prior income was the more important variable, and therefore appeared first. The group of 417 agents who reported earning less than \$90 prior to their appointment averaged only \$235,000 in first year sales. The average production of the remainder of the group was \$281,000. None of the other variables in the study proved capable of further dividing the group of 417 agents. The remaining 1108 agents were again

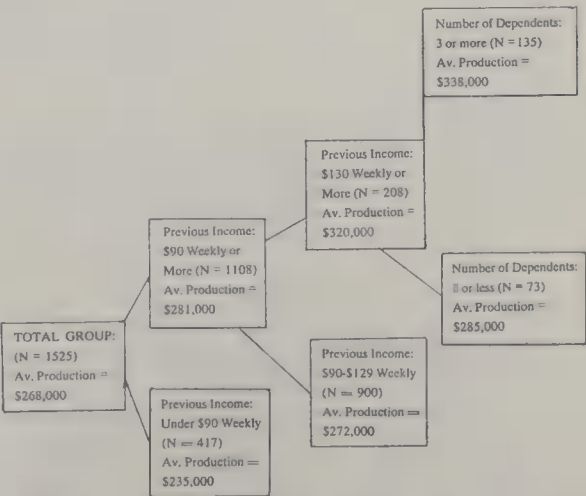


FIG. 1. Results of pattern analysis of biographical predictors of sales production (1963 appointees surviving at least 4 quarters).

TABLE 2

BIOGRAPHICAL ITEMS AND ITEM COMBINATIONS MOST HIGHLY RELATED TO SALES

| Rank | Group | Characteristic(s) | Average sales | SD | N |
|------|-------|--|---------------|-----------|------|
| 1 | 1 | More than \$130 prior weekly income and three or more dependents | \$338,000 | \$156,000 | 135 |
| 2 | 2 | More than \$130 prior weekly income and two or less dependents | 285,000 | 97,000 | 73 |
| 3 | 3 | Prior weekly income between \$90 and \$129 | 272,000 | 115,000 | 900 |
| 4 | 4 | Prior income under \$90 weekly | 235,000 | 96,000 | 417 |
| | | Total group | 268,000 | 117,000 | 1525 |

TABLE 3

VALUES OF STUDENT'S *t* FOR AVERAGE PRODUCTION OF GROUPS IDENTIFIED THROUGH PATTERN ANALYSIS

| Group | Group 1 | Group 2 | Group 3 | Group 4 |
|--|---------|---------|---------|---------|
| 1. More than \$130 prior weekly and 3 or more dependents | — | 2.13* | 3.82*** | 5.88*** |
| 2. More than \$130 prior weekly and 2 or less dependents | — | — | 0.87 | 3.30** |
| 3. Prior income between \$90 and \$129 weekly | — | — | — | 4.62*** |
| 4. Prior weekly income less than \$90 | — | — | — | — |

* *p* < .05, two-tailed test.
** *p* < .01, two-tailed test.
*** *p* < .001, two-tailed test.

partitioned by income into a group of 900 agents who averaged \$272,000 and a group of 208 whose sales averaged \$320,000. The group of 208 agents was then split according to number of dependents. An extremely high producing group of 135 agents emerged, along with a more moderately producing group of 73 agents. The large group of 900 agents could not be further divided by any of the remaining variables. The two groups formed on the basis of number of dependents were also not divisible.

The picture of the high producer given by this analysis is that of a man whose reported income prior to his appointment as an agent

was in excess of \$130 per week. This individual also reported on his application blank at least three dependents. The program identified the typical man who reported a prior salary of under \$90 a week as a low producer.

Table 2 summarizes the key characteristics of the four groups found by the pattern analysis program. The result of a *t* test of the production differences between these four groups is given in Table 3. The computer program appears, for the most part, to have succeeded in identifying groups reliably different in first year sales production. The analysis of variance reported in Table 4 suggests that the program also was successful in accounting for a significant portion of the total sum of squares.

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE OF PRODUCTION OF GROUPS IDENTIFIED BY COMPUTER PROGRAM FOR PATTERN ANALYSIS

| Source of variation | <i>df</i> | <i>M</i> _{Diff} | <i>F</i> |
|---------------------|-----------|--------------------------|----------|
| Between groups | 3 | 388,241 | 2.98* |
| Within groups | 1521 | 130,065 | |

* *p* < .05.

DISCUSSION

The general hypothesis of this study was confirmed. Distinct patterns were found. These patterns accounted for a significant proportion of the criterion variance.

Pattern analysis appears, on the basis of the authors' experience, to be a promising technique for the analysis of the kind of bio-

graphical variables of interest. If the findings survive cross-validation in a new sample they will be important considerations in recruiting and selection.

The results of the study are in agreement with the literature in that prior salary and number of dependents were again found to be associated with the sales production of life insurance agents. The present investigation added the information that prior income and number of dependents were sufficiently important in accounting for criterion variance to make superfluous any contribution of age, education, marital status, or previous sales experience.

No cross-validation was attempted. Since the program required groups to be of appreciable size before partitioning could occur, the entire sample was needed for the original analysis. It was believed that the small hold-out group that could have been spared would not reproduce the findings of the major sample. A small sample would not provide groups of the required size. Later agent groups will be used for the purpose of cross-validation. The large number of Ss remaining in certain of the groups, most notably the group of 900 agents, makes it desirable to add more biographical items to future analyses in order to

achieve finer groupings, and hence, more sensitive prediction.

REFERENCES

- DUNNETTE, M. D. *Personnel selection and placement*. Belmont, Calif.: Wadsworth, 1966.
- FERGUSON, L. W. Ability, interest, and aptitude. *Journal of Applied Psychology*, 1960, **44**, 126-131.
- LIFE INSURANCE AGENCY MANAGEMENT ASSOCIATION. *As they come to the job*. Hartford: Research Report No. 1954-4, 1954.
- LYKKEN, D. T., & ROSE, R. Psychological prediction from actuarial tables. *Journal of Clinical Psychology*, 1963, **19**, 139-151.
- MEEHL, P. E. *Minnesota Multiphasic Personality Inventory research for counsellors*. St. Paul, Minn.: St. Paul Department of Education, 1950.
- SHEPPS, R. R., TANOFKY, R., & MEAD, H. Life history information and success as an insurance salesman: The use of a "localized" salary history. New York: Metropolitan Life Insurance Company, 1967.
- SINES, J. O. Actuarial methods in personality assessment. In B. A. Maher (Ed.), *Progress in experimental personality research*. Vol. 3. New York: Academic Press, 1966.
- SONQUIST, J. A., & MORGAN, J. N. *The detection of interaction effects*. (University of Michigan Survey Research Center Monograph, No. 35) Ann Arbor: University of Michigan Press, 1964.
- SORENSEN, W. W. Configural scoring of biographical items for predicting sales success. Unpublished doctoral dissertation, University of Minnesota, 1964.

(Received March 20, 1968)

BRAND AWARENESS: DIFFERENTIAL ROLES OF FITTINGNESS AND MEANINGFULNESS OF BRAND NAMES¹

RABINDRA N. KANUNGO²

Dalhousie University, Halifax

Three experiments were performed to assess the comparative influence of fittingness and meaningfulness on the two stages of brand awareness: the response learning of brand names and the learning of brand-product association. The use of free-recall method in Experiments I and II revealed that only meaningfulness and not fittingness of brand names influences the response learning stage. Response learning was better for high meaningful than for low meaningful brand names. In Experiment III, an associative matching task was used to study the effects of meaningfulness and fittingness on the associative learning stage. The results suggest that while both the variables influenced the learning of brand-product association, it was fittingness of brand names that served as a better predictor of associative learning. Associative matching was better for the fitting than for the nonfitting brand names under both high and low meaningfulness conditions. Only in the restricted case of nonfitting brand names did meaningfulness influence associative learning.

Recent studies (Kanungo, 1968; Kanungo & Dutta, 1966) have shown that the meaningfulness and fittingness of brand names are potent variables influencing brand awareness. In these studies, the meaningfulness of a brand name was determined by the mean rating of the brand name by a group of Ss on a *meaningfulness scale* ranging from most to least meaningful categories. The fittingness of a brand name was determined by the degree of resemblance (in form, sound, and meaning) the brand name had to a common associate of the product which the brand name represented (Kanungo, 1968). With these operational measures for the two variables, the series of three experiments reported below was conducted to determine clearly differential roles of meaningfulness and fittingness on brand awareness.

It has been pointed out earlier (Kanungo & Dutta, 1966) that an advertisement of a product essentially contains a pair of items:

the product and the brand name. It is the purpose of the advertiser to make the consumer learn and retain the association between this pair of items. The situation, therefore, is analogous to paired-associate learning, which involves two separate stages: response learning and association learning (Underwood & Schulz, 1960). In order to learn pairs of items, the learner must first acquire the individual items themselves (response or free-recall learning), and then he must acquire an association between the pairs, so that given one of the items of the pair he can readily recall the other item (association learning or associative hookup). In testing for brand awareness, therefore, these two stages must be distinguished. It is conceivable that a consumer may recall a brand name, but may not be able to recall the product it represents. In this case, the consumer shows evidence for response learning but shows no evidence for having learned the brand-product association.

Earlier studies (Murdock, 1960; Underwood & Schulz, 1960) have suggested that response learning of words is linearly related to their frequency of usage or familiarity. The attribute of meaningfulness has been shown to covary with familiarity (Noble, 1953, 1954). Thus it may be hypothesized that learning and recall of brand names alone

¹ The study was supported partly by Grant No. X-12-179 from the National Research Council of Canada and partly by Grant No. X-84-124 from Dalhousie University Research Development Fund. The author is grateful to Marcia Earhard for many helpful comments.

² Requests for reprints should be sent to the author, Department of Psychology, Dalhousie University, Halifax, Nova Scotia.

(without reference to the product they represent) will be directly determined by the meaningfulness of the brand names. The fittingness variable, on the other hand, should have direct influence on the acquisition of brand-product association because, by definition, the fitting brand names have closer resemblance with existing common associations to the product than the nonfitting brand names.

EXPERIMENT I

Experiment I was designed to study the effect of the meaningfulness variable on the response learning (or free recall) of the brand names. The brand names were presented to Ss in such a manner that the brand names appeared to vary only in their meaningfulness (high or low) but not in their fittingness characteristics.

Subjects

Thirty-two undergraduate students (16 males and 16 females) served as Ss. These Ss had not participated in any kind of psychological experiment before.

Materials and Procedure

All the 48 brand names selected and reported in an earlier study (Kanungo, 1968) were used in the present study. These brand names had been selected on the basis of their meaningfulness and fittingness for 12 different products. Thus for each product there were four brand names: high meaningful-fitting (HM-F), low meaningful-fitting (LM-F), high meaningful-nonfitting (HM-NF), and low meaningful-nonfitting (LM-NF).

Four lists, each containing a different set of 12 brand names, were prepared from the total 48 brand names for the purpose of this experiment. Each list contained only the brand names of each of the 12 products in such a manner as to include 3 HM-F, 3 HM-NF, 3 LM-F, and 3 LM-NF brand names. Eight Ss were randomly assigned to each list. The Ss were made to believe that the words contained in the list were real brand names used by some advertisers for their products. They were told that as a part of an advertising research *E* was interested in knowing the reactions of Ss to each of the brand names.

Each *S* was given a list of the brand names in random order without any mention of the products the brand names might represent. The *S* was asked to rate each brand name for its meaningfulness and its appropriateness (or suitability) for being used in the advertisements. Twelve answer sheets were provided to *S*, one for each brand name, to record his ratings on the two printed 6-point scales of

meaningfulness and appropriateness. The 6 points on the scales were labeled as follows: very high, high, moderate, low, very low, none. The *S* was instructed to rate the brand names by underlining the appropriate verbal label. The order of presentation of the two scales was counterbalanced for each *S*. To ensure that Ss paid proper attention to the brand names in the list, each *S* was asked to write down the brand name at the top of the rating sheet.

Each *S* was given approximately 1 min. to rate each brand name on the two scales. After *S* had finished rating all the brand names, the list and the answer sheets were collected. Then *S* was given 5 min. to recall and write down as many brand names from the list as possible.

Results and Discussion

The Ss' ratings of the brand names on the two scales were transferred to ordinal weights ranging from 5 (very high) to 0 (none) for the 6 points on the rating scales. Each brand name was rated by 8 Ss. Taking the mean ratings for each brand name, it was observed that the 24 HM brand names had significantly higher mean meaningfulness values (mean $m' = 3.62$) than the 24 LM brand names (mean $m' = 1.60$, $t = 5.66$, $p < .001$). This suggests that Ss did perceive the HM and LM brand names selected by *E* as differing in meaningfulness. The appropriateness ratings revealed that the F brand names were considered no more appropriate (mean appropriateness = 2.62) than the NF brand names (mean appropriateness = 2.57, $t = 0.10$). This result was expected, in view of the fact that the fittingness of the brand names for specific products was not revealed to Ss. The brand names had been presented as discrete words without the context of the product they represented. This indicated that the experimental attempt to manipulate meaningfulness apart from fittingness was successful.

Effect of meaningfulness. An analysis of variance was performed on the free-recall scores using Treatments \times Treatments \times Subjects design (Lindquist, 1953). The two treatment classifications were HM versus LM and F versus NF brand names. Only one main treatment effect, HM versus LM brand names, was significant ($F = 15.41$, $df = 1/31$, $p < .01$). The mean recall scores for HM-F and LM-F brand names were 1.81 and 1.19, respectively ($t = 2.95$, $p < .01$). The mean re-

call scores for HM-NF and LM-NF brand names were 1.78 and 1.22, respectively ($t = 2.67, p < .01$). These results support the contention that the response learning of brand names is directly related to the meaningfulness of the brand names. The HM brand names were better recalled than LM brand names in a free-recall situation.

EXPERIMENT II

In order to highlight the isolated effects of the meaningfulness variable on the response learning stage of brand awareness, the free-recall test in Experiment I was administered after the brand names had been presented as words in a list without the context of the products they represented. In practice, however, brand names always appear in the context of the products they represent. Thus when a brand name appears in an advertisement, its meaningfulness as well as its fittingness for the product become salient for the consumer, and the two variables simultaneously influence brand awareness (Kanungo, 1968). It was not clear from the result of Experiment I whether only meaningfulness and not fittingness influences response learning of brand names in real advertisements where both variables are salient. Experiment II was designed to answer this question.

In addition, Experiment II attempted to replicate the findings of Kanungo (1968) regarding the effect of the utility of the product a brand name represents on brand-name recall. Unlike the results reported by Kanungo and Dutta (1966) using Ss from India, Kanungo (1968) reported that the brand names for male-use and female-use products were recalled equally well by both male and female Canadian Ss. However, it was noticed that both groups recalled fewer brand names of products used by both males and females, perhaps due to the fact that these products, such as adhesive tape or folders, had less personal significance or importance for Ss.

Subjects

Twenty-four adult male and 24 adult female undergraduate students served as Ss. These Ss had not participated in any kind of psychological experiment before.

Materials

The same materials used and described in detail by Kanungo (1968) were used in this experiment. Briefly, four commonly used products from each of the three product-utility categories (male-use, female-use, and used-by-both products) were chosen. For each of the 12 products, four brand names, a HM-F, a LM-F, a HM-NF, and a LM-NF were selected. Using these brand names, four advertisements for each product were prepared. Twelve copies of each of the 48 advertisements were printed on a 14-cm. \times 22-cm. art paper. In each printed layout, the brand name appeared at the top of a picture of the product. Below the picture, a short phrase containing the product name was present. Farther down the page, the name and address of the advertiser (a fictitious manufacturing concern) were given.

Using these advertisements, 48 booklets were compiled, each containing an advertisement of each of the 12 products. Each booklet included a HM-F, a LM-F, a HM-NF, and a LM-NF brand name for each of the three product categories—male-use, female-use, and used-by-both products. The sequence of the 12 advertisements in each booklet was randomized to ensure varied order of presentation to Ss.

Procedure

Each S was given a booklet and was asked to rate each brand name contained in the booklet on each of two 6-point scales: meaningfulness and appropriateness (cf. Kanungo, 1968). The Ss were made to believe that the advertisements were genuine and that the advertiser wanted to know the reactions of Ss to the brand names in terms of the two scales before launching an extensive advertising campaign.

The Ss were given approximately 1–1.5 min. to rate each brand name on the two scales. After Ss had finished rating all the brand names, the booklets were collected from them. Then Ss were given blank sheets of paper for the free recall of brand names. They were asked to write down on these sheets the brand names from the booklets that they could recall within a maximum period of 5 min. The free-recall test, instead of an aided-recall test where the product names are given to Ss (Kanungo, 1968), was administered because the former is a more sensitive method of assessing response learning of brand names.

Results and Discussion

Ratings of brand names. The ratings of the brand names on the two scales were analyzed in the same manner as in Experiment I. In the present experiment each brand name was rated by 12 Ss. The Ss' ratings indicated that the 24 HM brand names were significantly higher in meaningfulness (mean $m' = 2.92$) than the 24 LM brand names (mean $m' = 1.99, t = 3.44, p < .01$). The appropriateness

ratings revealed that the F brand names were considered significantly more appropriate (mean appropriateness value = 2.81) than the NF brand names (mean appropriateness value = 1.71, $t = 4.07$, $p < .001$). These ratings are very similar to those reported by Kanungo (1968) and suggest the validity of the experimental manipulations of meaningfulness and fittingness of the brand names.

Effects of fittingness and meaningfulness. A $2 \times 2 \times 2$ analysis of variance (Lindquist, 1953, Type VI design) was performed on the free recall of brand names to assess the relative influence of meaningfulness and fittingness variables. The three classifications were meaningfulness (HM versus LM) and fittingness (F versus NF) of brand names, and Ss' sex (male versus female). Only the main effect of meaningfulness was found to be significant ($F = 18.88$, $df = 1/46$, $p < .01$). The interaction effects were not significant. However, it may be pointed out that the interaction between meaningfulness and fittingness variables approached statistical significance ($F = 3.48$, $df = 1/46$, $.05 > p < .10$). The reason for this trend in the interaction effect can be seen from Table 1, where the mean recalls of HM-F, LM-F, HM-NF, and LM-NF brand names by all 48 Ss are presented. The F ratio for meaningfulness indicated that the free recall of the HM brand names is significantly better than those of LM brand names. However, the mean recall scores presented in Table 1 reveal that such significance stems mainly from the recall of NF brand names. Thus when the brand names

are of the nonfitting type, the effects of meaningfulness on free recall of brand names are more clearly observed ($t = 2.65$, $p < .01$). It will also be noticed that the free recall of F brand names does not differ significantly from those of NF brand names.

These findings are unlike those reported by Kanungo (1968), who used aided-recall procedure, in which Ss were presented with the product names and then were asked to recall appropriate brand names for the products. Except for the recall procedures, the materials and the method used in both the Kanungo (1968) study and the present one are very similar. The mean recall scores reported by Kanungo (1968) are also presented in Table 1, for the purpose of comparison. It will be seen from Table 1 that the potential influence of fittingness on brand awareness as shown by Kanungo (1968) disappeared in the present study where the free-recall test was used. Since the free-recall test is a test of brand-name learning, and not of the learning of brand-product association, it may be concluded that it is the meaningfulness and not fittingness of the brand names that determines brand-name learning.

Effect of product utility. The mean free-recall scores of the male and female Ss for the brand names of the three categories, male-use, female-use, and used-by-both products, are presented in Table 2. A 2×3 analysis of variance performed on the free-recall scores revealed that only the main effect of the three product categories was significant ($F = 4.00$, $df = 2/92$, $p < .025$). The reason for the

TABLE 1
MEAN RECALL AND MATCHING SCORES FOR BRAND NAMES

| | Kanungo (1968) study: Aided recall ^a | | | Experiment II: Free recall ^a | | | Experiment III: Matching ^b | | |
|-----------------|--|------------|----------|--|------------|----------|--|------------|----------|
| | Fitting | Nonfitting | <i>t</i> | Fitting | Nonfitting | <i>t</i> | Fitting | Nonfitting | <i>t</i> |
| High meaningful | 2.81 | 2.46 | 2.33* | 2.11 | 2.15 | 0.15 | 11.00 | 9.67 | 4.35*** |
| Low meaningful | 2.56 | 1.50 | 7.07*** | 1.90 | 1.46 | 1.69 | 10.63 | 6.00 | 15.13*** |
| <i>t</i> | 1.67 | 6.40*** | | 0.81 | 2.65** | | 1.21 | 11.99*** | |

^a $N = 48$.
^b $N = 24$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

significant effect of the product-utility variable stems from the fact that Ss recalled relatively fewer brand names for the used-by-both products than for the other two categories (see Table 2).

These results can be compared with the results obtained by Kanungo (1968). Table 2 also presents the mean recall scores reported in the Kanungo (1968) study. Two observations can be made from such a comparison. First, the recall scores obtained by Kanungo (1968) are higher than those reported in the present experiment. The reason for this observation is quite obvious. Kanungo (1968) used aided-recall method in which Ss were presented with the product names which served as cues for recalling the appropriate brand names. In the present experiment, Ss were required to recall brand names without being given any product names to serve as cues for their recall. Second, the patterns of recall of brand names for the three product-utility categories in both the present and the earlier (Kanungo, 1968) experiments are very similar, despite the differences in their recall procedures. In both studies, there was a trend toward lower recall of brand names for the used-by-both products than for either the male-use or the female-use products (see Table 2). The lower recall of brand names for the used-by-both products might have resulted from the impersonal nature of the

products belonging to this category. For instance, products like adhesive tape, folder, or writing pads, used in both the experiments, presumably had less personal significance for the Ss, and hence they may have been considered less important than the products belonging to the other two categories such as nylons or men's shorts. Another additional factor that might have produced the lower brand name recall for the used-by-both products is exposure in real life to advertisements for the products used in this study. The male- and female-used products such as shirts, nylons, girdles, etc., are more widely and frequently advertised than products such as folders or adhesive tape. This might have made Ss more attentive and sensitive to the advertisements representing the former than the latter type of products. Whether it is the personal significance of the product for the consumer, or the frequency of exposure in real life to the advertisements of the product, or both these factors taken together that influence brand-name recall is a question that calls for further research.

EXPERIMENT III

The free-recall method used in Experiments I and II revealed the differential effects of meaningfulness and fittingness variables only on the response learning of brand names, but not on the learning of brand-product association. Thus the final experiment was designed to assess the comparative influence of the two variables on the learning of brand-product association. The learning of associations between the brand names and the products they represent was tested through an associative matching task. The Ss were provided with both the brand names and the product names after exposure to the advertisements, and were asked simply to pair or match the brand names with the products they represented in the advertisements. Such matching task provided a situation in which individual differences among Ss in their response learning of brand names was eliminated (since all the brand names were available to Ss), and Ss' matching performance unambiguously reflected their learning of brand-product associations.

TABLE 2
MEAN RECALL SCORES OF MALE AND FEMALE SUBJECTS
FOR THE THREE CATEGORIES OF BRAND NAMES

| Study | Product | | |
|------------------------------|----------|------------|--------------|
| | Male use | Female use | Used by both |
| Experiment II (free recall) | | | |
| Male ^a | 2.79 | 2.50 | 2.08 |
| Female ^a | 2.62 | 2.83 | 2.42 |
| Combined ^b | 2.71 | 2.67 | 2.25 |
| Kanungo, 1968 (aided recall) | | | |
| Male ^a | 3.29 | 3.12 | 2.87 |
| Female ^a | 3.17 | 3.25 | 3.00 |
| Combined ^b | 3.23 | 3.19 | 2.94 |

^a N = 24.
^b N = 48.

Subjects

Twelve male and 12 female undergraduate students served as Ss. These Ss had never participated in any kind of psychological experiment before.

Materials

The same advertisements used in Experiment II were also used in this experiment. This time, however, when the advertisement booklets were compiled, each booklet included 48 advertisements: four advertisements of each of the 12 products. The four advertisements for each product included one HM-F, HM-NF, LM-F, and LM-NF brand name. The sequence of the 48 advertisements in each booklet was randomized to ensure varied order of presentation to Ss.

Procedure

Each S was given a booklet and was asked to rate each brand name contained in the booklet on each of the two 6-point scales: meaningfulness and appropriateness, in a manner similar to Experiment II. After the ratings were completed, Ss were given a single serial list of brand names and product names listed in random order. This procedure ensured that all the responses were available to each S, and it also controlled for differences in the response learning of brand names among Ss. The Ss were asked to use the list to pair each brand name with the product name it represented in the advertisements and to write down the pairs on a separate sheet of paper. A maximum of 10 min. was allowed for this associative matching task.

Results

Rating of brand names. Each brand name was rated on each of the two scales by 24 Ss. These ratings were analyzed in the same manner as described for Experiments I and II. The mean m' values for HM and LM brand names were 2.55 and 1.77, respectively ($t = 2.89$, $p < .01$). The mean appropriateness values for F and NF brand names were 2.63 and 1.67, respectively ($t = 4.30$, $p < .001$). These results are very similar to those of Experiment II and they again substantiate the validity of the experimental manipulations of the two variables: meaningfulness and fittingness.

Effects of meaningfulness and fittingness. In order to assess the relative influence of meaningfulness and fittingness on the correct matching of brand names with the product names, a $2 \times 2 \times 2$ analysis of variance Type VI design (Lindquist, 1953) was performed on the number of correct matchings made

by Ss. Again the three classifications were meaningfulness (HM versus LM brand names), fittingness (F versus NF brand names), and Ss' sex. Both the meaningfulness and fittingness variables had significant effects ($F_s = 74.25$ and 142.01 , respectively, with $df = 1/22$ and $p < .001$ in each case). The matching scores of male and female Ss did not differ significantly ($F = 2.14$, $df = 1/11$, $p > .05$). Two of the interaction effects were also found to be significant. The F ratios for the interaction between meaningfulness and fittingness was 56.53 ($df = 1/22$, $p < .001$) and between Ss' sex and fittingness was 5.84 ($df = 1/22$, $p < .05$). The mean matching scores of all 24 Ss for HM-F, LM-F, HM-NF, and LM-NF brand names are presented in Table 1. These means reveal the reason for the significant interaction between meaningfulness and fittingness. Associative matching was significantly better for HM than for LM brand names only when the brand names were nonfitting ($t = 11.99$, $p < .001$). However associative matching was better for F than for NF brand names regardless of the meaningfulness of the brand names (see Table 1), suggesting that fittingness of a brand name may be a better predictor of learning of brand-product association.

A comparison between the matching scores of male and female Ss revealed the reason for the significant interaction between Ss' sex and fittingness. The difference between the scores of male Ss ($M = 13.83$) and female Ss ($M = 17.50$) for NF brand names was significant beyond the .001 level ($t = 7.34$), but for the F brand names the difference between the scores of male Ss ($M = 21.00$) and female Ss ($M = 22.25$) was significant only at the .05 level ($t = 2.50$).

DISCUSSION

The results of the three experiments clearly demonstrated the differential roles of meaningfulness and fittingness of brand names on the two stages of brand awareness: learning of brand names and learning of brand-product association. The results of Experiments I and II revealed that the major determinant of learning and recall of a brand name (regardless of the product context in which it appears) is its meaningfulness and

not its fittingness characteristic. The results of the third experiment suggested that while both meaningfulness and fittingness of the brand names may influence the learning of brand-product association, it is the latter variable that seems to serve as a better predictor of the associative learning. Under the conditions of both high and low meaningfulness, the fittingness characteristics of the brand names favored their associative learning. Only in the restricted case of nonfitting brand names did meaningfulness influence learning of brand-product association (see Table 1). Among nonfitting brand names, the brand-product association was formed faster if the brand names were of high meaningfulness than if they were of low meaningfulness. This finding is consistent with the associative probability notion proposed by Underwood and Schulz (1960) for paired-associate learning. A highly meaningful brand name evokes a larger number of associations and hence there is a greater possibility that it will find a

common link or basis of association with the product sooner than a low meaningful brand name evoking fewer associations.

REFERENCES

- KANUNGO, R. N. Brand awareness: Effects of fittingness, meaningfulness, and product utility. *Journal of Applied Psychology*, 1968, **52**, 290-295.
- KANUNGO, R. N., & DUTTA, S. Brand awareness as a function of its meaningfulness, sequential position, and product utility. *Journal of Applied Psychology*, 1966, **50**, 220-224.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- MURDOCK, B. B., JR. The immediate retention of unrelated words. *Journal of Experimental Psychology*, 1960, **60**, 222-234.
- NOBLE, C. E. The meaning-familiarity relationship. *Psychological Review*, 1953, **60**, 89-98.
- NOBLE, C. E. The familiarity-frequency relationship. *Journal of Experimental Psychology*, 1954, **47**, 13-16.
- UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Philadelphia: Lippincott, 1960.

(Received March 21, 1968)

INTERACTION OF ACHIEVEMENT CUES AND FACILITATING ANXIETY IN THE ACHIEVEMENT OF WOMEN¹

W. J. McKEACHIE²

University of Michigan

In three studies, 250-380 women psychology students low in facilitating anxiety (FA) as measured by the Alpert-Haber AAT achieved better grades when taught by teachers characterized by expectations of high standards of achievement than when by teachers not so characterized, while high FA women performed more poorly in classes taught with high standards of achievement.

Alpert and Haber (1960) developed a test of anxiety about achievement tests with two subtests—one for debilitating anxiety, the other for facilitating anxiety. They presented evidence that the Facilitating Anxiety Scale correlated positively with grade-point average (GPA) while the Debilitating Anxiety Scale correlated negatively with GPA.

Use of the Alpert and Haber test at the University of Michigan was found to contribute little beyond standard college aptitude measures to the prediction of grades in the introductory psychology course. Scores on the scale, however, were found to interact consistently with a measure of achievement cues emitted by the instructor in predicting the grades of women (not men). The sample and procedures used in these studies are described by McKeachie, Lin, Milholland, and Isaacson (1966).

Table 1 indicates this interaction for three different samples. In each sample women students high in facilitating anxiety did relatively well in classes low in achievement cues while students low in facilitating anxiety did relatively poorly in these classes.

To account for this rather consistent finding let us look first at the specific measures of achievement cues and facilitating anxiety. In the 1961 and 1963 studies the index of

achievement cues was obtained by computing the mean rating which students assigned an instructor on the item, "He maintained definite standards of student performance." In the 1958 study McKeachie et al. used the mean of three items: "Instructor set very high standards for students," "Members of the class competed to do well," and "The course work presented a real challenge to me."

A typical item on the Facilitating Anxiety Scale is "I work most effectively under pressure—as when the task is very important." Students high in facilitating anxiety (who answered such questions affirmatively)

TABLE 1
INTERACTION OF FACILITATING ANXIETY,
ACHIEVEMENT CUES, AND GRADES
OF WOMEN

| Study | Facilitating anxiety | Achievement cues | Grades | | Total |
|-------|----------------------|------------------|--------|-----------|-------|
| | | | A & B | C, D, & E | |
| 1963 | Hi | Hi | 27 | 25 | 52 |
| | | Lo | 40 | 27 | 67 |
| | Lo | Hi | 37 | 23 | 60 |
| | | Lo | 20 | 38 | 58 |
| Total | | | 124 | 113 | 237 |
| 1961 | Hi | Hi | 55 | 47 | 102 |
| | | Lo | 54 | 42 | 96 |
| | Lo | Hi | 35 | 40 | 75 |
| | | Lo | 40 | 65 | 105 |
| Total | | | 184 | 194 | 378 |
| 1958 | Hi | Hi | 50 | 36 | 86 |
| | | Lo | 72 | 49 | 121 |
| | Lo | Hi | 40 | 43 | 83 |
| | | Lo | 41 | 49 | 90 |
| Total | | | 203 | 177 | 380 |

Note.—Interaction chi-square for combined data = 5.59 with 1 df; $p < .02$.

¹ The earlier data reported in this study were collected under a grant from the Fund for Advancement of Education. The later data were collected and analyzed with support from the United States Office of Education, Research Contracts SAE-8541 and 4/01-001 to W. J. McKeachie, J. E. Milholland, and Robert L. Isaacson. Yi-Guang Lin carried out the data analysis.

² Requests for reprints should be sent to the author, Department of Psychology, University of Michigan, Ann Arbor, Michigan 48104.

were expected to do well in classes with high achievement cues. Results in the opposite direction suggest that the Facilitating Anxiety Scale may be a test of general academic motivation as much as a scale of test anxiety. In short, women who score high on this scale may say they are unafraid of tests because they study and are prepared to the limits of their ability. Students who score low on the scale may simply be unmotivated. When the instructor provides additional cues to achievement, the unmotivated students begin to work. Since grades are relative within a class, the unmotivated students win more of the high grades. This suggestion is supported by the finding that on the Criteria Test of Psychological Thinking (Milholland, 1964) and on a test of knowledge administered as part of the final examination, women high in facilitating anxiety did about equally well whether taught

in high or low achievement cue sections, but the low facilitating anxiety women more nearly approached the achievement of the high facilitating anxiety women when in high cue sections.

REFERENCES

- ALPERT, R., & HABER, R. N. Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology*, 1960, **61**, 207-215.
- McKEACHIE, W. J., LIN, Y., MILHOLLAND, J., & ISAACSON, R. Student affiliation motives, teacher warmth, and academic achievement. *Journal of Personality and Social Psychology*, 1966, **4**, 457-461.
- MILHOLLAND, J. E. Measuring cognitive abilities. In W. McKeachie, R. Isaacson, & J. Milholland (Eds.), *Research in characteristics of effective college teaching*. Ann Arbor: University of Michigan, 1964.

(Received March 23, 1968)

RELATION BETWEEN BIRTH ORDER AND BEING A BEAUTICIAN

PHILIP S. VERY¹ AND JOSEPH A. ZANNINI

Rhode Island College

The basic premise of the study is that birth order as described by Konig is a factor which greatly influences personality, and that personality in turn is an important factor in vocational choice. Konig describes the second born as being easygoing, seeking harmony and the like; beauticians ($N = 210$) were chosen as a vocation typifying these traits. Of the 210 Ss, a statistically significant number were second born.

A survey of the literature on birth order and personality indicates that while differences do exist, extensive investigations and clarification are needed. An excellent review of the literature on birth order and its results has been presented by Altus (1966). Altus' text contains summaries of studies by Galton, Ellis, Clarke, Apperly, Jones, and Roe, who all found that there was a significantly greater number of firstborn in eminent positions than those born later.

Much less work has been done on personality than on intelligence or success primarily because of the difficulties of proper measurement. The "grand old man" of birth order and personality research is, of course, Alfred Adler. Using clinical study rather than empirical research, Adler postulated a natural rivalry arising between siblings, the firstborn being "born into" a prominence in the family, the second born and others attempting to "overthrow" his position. Feelings of inferiority generating a drive toward superiority were the essential motivating factors in his schema. What must be remembered was that Adler practiced in Vienna during the Victorian period and saw a clientele which consisted of many highly repressed, sometimes persecuted, overly ambitious people. The picture changes somewhat in America, which presents to the new-born a more liberal environment and certainly one in which the laws of primogeniture are not nearly so significant as they were in Europe during the past half-century.

In his book, *Brothers and Sisters*, Karl Konig (1963) states that there is a definite

relationship between personality and birth order. Konig's theory states that there are four basic sibling positions; only child, first-born, second born, and third born. The triadic pattern of first, second, and third born repeats itself through succeeding siblings. The firstborn is described as a defender of the family's attitudes, being socially responsible, somewhat domineering, ambitious, aggressive, independent, and a leader. The second born is more casual, leisurely, and harmonious. His basic personality pattern does not contain the attributes that one expects a leader or a defender to possess. The third born is described as overly sensitive, withdrawn, reflective, and usually feels lonely and segregated. The only child is a combination of the first and third born personality traits.

The purpose of this study is to discover whether or not there is a relation between birth order, as described by Konig, and being a beautician. It is hypothesized that second borns will be present in a random sample of female beauticians statistically significantly greater than expected by chance. Beauticians were chosen to be subjects in this study, for it was assumed that a beautician would typify the personality traits of a second born. This assumption was borne out by both instructors and managers of beauty salons, for they all emphasized such second born traits as courtesy, cooperativeness, and self-control. Availability and ease in which one could contact beauticians was also a factor in selecting beauticians for the study.

METHOD

Subject. The total sample consisted of 210 female beauticians from the state of Rhode Island. The

¹ Requests for reprints should be sent to Philip S. Very, Rhode Island College, 600 Mount Pleasant Avenue, Providence, Rhode Island 02908.

beauty salons were selected by numbering two lists of salons from the local telephone directory (one urban, the other rural). A table of random numbers was used to select 60 beauty salons from the total listings (85% urban and 15% rural, corresponding to the population of the state).

The mean age of Ss was 25.35 yr. and the standard deviation was 6.15.

Questionnaire. The questionnaire used required Ss to indicate their own age, sibling ranking, and sex, plus those of their brothers and sisters. The questionnaire also asked for the year of death of any sibling. The questionnaire indicated that only siblings born alive were to be reported.

Procedure. The data were obtained simply by entering those salons selected and asking those beauticians there to fill out the questionnaire.

The questionnaires were then filed into appropriate categories of only, first, second, and third born. The firstborn was defined as the first of two or more children and does not include only children which are a class in themselves. As stated above, there is a triadic pattern of first, second, and third born positions which repeats itself through succeeding siblings. Therefore, fourth borns were classified as firstborns, fifth borns were classified as second borns, etc.

RESULTS

Of the 210 Ss, 15 were only children, 74 were firstborns, 87 were second borns, and 34 were third borns, yielding 7.14%, 35.25%, 41.43%, and 16.19%, respectively. In order to obtain an expected proportion of second-born females that would be appropriate, the data on second-born females, born from 1924 to 1948 (according to the National Office of Vital Statistics, 1959), were used.

Since size of family is affected by such external environmental factors as depressions, wars, and the like, it was decided to compute what may be called a weighted average of the census percentages for second borns. Ordinary averages over 5-yr. periods from 1924 through 1948 were computed; then these averages were weighted by the numbers in the actual sample born in those 5-yr. periods in order to arrive at an appropriate expected percentage of second borns for theoretical comparisons. The final weighted average expected value was 33.09%.

The possible significance of the larger number of second borns was tested by using the significance of the difference of a sample percentage from a theoretical value. This yielded a standard deviation of 3.16%; therefore,

41.43% is 2.64 standard deviations above the expected value of 33.09 and is thus significant at the .01 level.

DISCUSSION

Although the results obtained are statistically significant, there are many interacting variables which seem to confound the results and reduce the possibility of even greater significance. It was indirectly observed, while distributing the questionnaires, that there are different motivating factors, which lead a female to enter the field of cosmetology. This is especially true since it takes relatively little formal training and intellectual ability to be a licensed beautician.

Some beauticians have expressed the fact that they became beauticians because they felt that with a license they could, when necessary, work at home and still raise a family. There are others who are totally engulfed with the idea of being artistic and creative as beauticians. And there are status seekers who try to capitalize on the artistic end of the field and attempt to project an image of the beautician as a highly trained and skilled professionalist. It may be noted here that although the state has only one license for beauticians, there are different occupational titles, such as cosmetologist, hair stylist, and beautician, which are usually only used in a status sense. In essence, these different attitudes reflect different birth order personalities.

In spite of the overwhelming evidence favoring the success of the firstborn, virtually nothing has been done in the United States to try to provide environmental compensation to the ordinal positions. That a child should be seriously affected in his life's goals as a result of his order of birth is a possibility abhorrent to humanitarian America. Nevertheless, the weight of research is beginning to bear this out. What is needed is continued clarification of the reasons for this achievement differential. One of the most promising avenues is, of course, the study of the personality differences and motivation which are at least partially responsible for the differ-

ential success. Only then can some remedial measure be taken to offset this factor.

REFERENCES

ALTUS, W. D. Birth order and its sequelae. *Science*, 1966, 15, 44-48.

KONIG, K. *Brothers and sisters*. New York: St. George Press, 1963.

NATIONAL OFFICE OF VITAL STATISTICS. *Vital Statistics—Special Reports, National Summaries*. 50, November 17, September 14, 1959.

(Received March 23, 1968)

(Continued from page 106)

Interference between Concurrent Tasks of Driving and Telephoning: I. D. Brown,* A. H. Tickner, and D. C. V. Simmonds: Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

Comparison of Several Patterns of Communication: Charles J. O'Leary and Morton Goldman*: Psychology Department, University of Missouri, Kansas City, Missouri 64110.

Comparison of 'Real Life' Assessments of Management Potential with Situational Exercises, Paper and Pencil Ability Tests, and Personality Inventories: J. R. Hinrichs*: IBM Corporation, 112 East Post Road, White Plains, New York 10601.

Image of Industrial Psychology among Personnel Administrators: George C. Thornton, III*: Department of Psychology, Colorado State University, Fort Collins, Colorado 80521.

Work Values and Job Satisfaction: Milton R. Blood*: Department of Psychology, University of California, Berkeley, California 94720.

Effects of Performance on Leadership, Cohesiveness, Influence, Satisfaction, and Subsequent Performance: George F. Farris* and Francis G. Lim, Jr.: Massachusetts Institute of Technology, Sloan School of Management, Room 52-590, 50 Memorial Drive, Cambridge, Massachusetts 02139.

* Asterisk indicates author for whom address is supplied.

PSYCHOLOGICAL CONCOMITANTS AND DETERMINANTS OF VOCATIONAL CHOICE¹

KENNETH M. KUNERT²

University of California, Berkeley

The Vocational Life Patterns (VLP) *Q* Sort was developed and used to investigate the personality-vocational-choice relationship. Groups of 75 student Ss, 50 Ss for the initial study and 25 Ss for the cross-validation study, were drawn from the schools of law, medicine, theology, and engineering. Results of the student *Q*-sort distributions confirm that the motivational and personality self-concept descriptions of the VLP *Q* Sort do distinguish these different vocational groups at highly significant levels. Descriptions are presented of the different personality-vocational patterns of the four groups. It was concluded that the VLP *Q* Sort was effective in investigating the personality-vocational-choice relationship and it was suggested that the procedure provides a means to study the more elusive constructs of personality theories.

For the counseling psychologist the relationship between personality and vocational choice is of much interest. Ginzberg, Ginsburg, Axelrad, and Herma (1951) discussed this personality-vocational-choice relationship in terms of the individual's values and goals. A little later, Darley and Hagenah (1955) envisioned the relationship as embracing one's needs, value systems, and motivations. Roe (1956) applied Maslow's theory of motivation in an attempt to unite theory and research findings to give evidence for this relationship. Super (1953, 1957) presented the well developed theory that vocational choice implements the self-concept.

The present research, stimulated by the above and similar ideas, has applied the *Q*-sort methodology to the question of the personality-vocational-choice relationship. This study was undertaken to determine whether distinctive personality patterns could be found for students committed to the fields of engineering, law, medicine, and theology.

¹ This article is based upon a doctoral dissertation submitted to the Graduate School of the University of California, Berkeley, in 1965. Appreciation is expressed to Harrison G. Gough, dissertation chairman, and also to Lyman W. Porter and Gordon H. Robinson, dissertation readers, for their help and suggestions. Thanks are also given to Quenten Welsh for his assistance in executing the computer programming and analysis.

² Now at the Department of Psychology, University of Detroit, Detroit, Michigan 48221. Requests for reprints should be sent to the author at the above address.

Recently, a few articles have reported using the *Q* sort in this area. Englander (1960, 1961) had used a variation of the *Q*-sort method and found that the self-concepts of elementary teaching trainees did seem to be implemented by their vocational choice. Morrison (1962) used the *Q*-sort method to find that the self-concepts of nursing trainees and of teacher trainees were related to their concepts of the nursing and teaching professions. In 1963, Neff and Helfand used a *Q* sort to study the meaning of work. They found that the work potential of 16 physically handicapped persons could be classified by comparing their individual sorts with a criterion sort developed in the study. All of these studies employed either a self- and an ideal-sort or a self-sort versus a criterion sort.

The present investigation, however, used only the self-concept sort in obtaining its data and deriving its results. The Ss described themselves through the Vocational Life Patterns (VLP) *Q*-sort deck. It was hypothesized that the different vocational fields being researched could be differentiated by the personality and motivational variables of the VLP *Q* sort.

METHOD

Subjects

The Ss were students obtained from the vocational fields of law, medicine, theology, and engineering. They were chosen from these fields since each field provides a clear and distinct vocational choice and at the same time the fields allow a sampling

of the interest continuum which ranges from humanitarian interests at one end to things/object interests at the other end. Also the selectivity factors involved in getting into these fields permits the making of the assumption that these Ss would show equal enthusiasm and commitment to their choices. Seventy-five Ss were sought in each of these professional fields with 50 Ss being randomly assigned to the original sample and the remaining 25 Ss being assigned to the cross-validation sample. The Ss for each sample were 21–29 yr. Graduate students were used in the engineering group so that all groups might be of comparable educational levels. The theology group combined students from Protestant and Catholic seminaries to control for differences of religious background and of marital status found in the other three groups. All Ss were obtained from professional schools in the San Francisco Bay Area.

Procedure

The Ss performed the VLP *Q* sort³ in a group setting. Each *S* was given the *Q*-sort deck containing a title card, seven category identification cards, and the 70-item statement cards. Each *S* was also given a combined instruction and recording sheet. The instructions explained the *Q*-sort procedure and defined the meaning of the seven categories which ranged from most characteristic to most uncharacteristic. The Ss were asked to: "Please read these statements and then classify them according to their importance and accuracy in describing yourself as you think of yourself at the present time." The recording part of the sheet contained columns for each of the seven categories, each column having the appropriate number of spaces in which to write the number of the statements placed in the particular category. The item statements were to be distributed across the seven categories as follows: 4—9—14—16—14—9—4. Items assigned to each category were given the weighted score for that category. The weighted scores ranged from 7 to 1 with 7 being assigned to the most characteristic category and 1 to the most uncharacteristic category and the weights of 6 to 2 being assigned to the categories in descending order insofar as the category was more or less characteristic in its designation.

The 70 items of the VLP *Q* sort express ideas representative of four major topics with 18 subtopic areas. A delineation of the topic and subtopic areas follows:

1. The recreational or leisure time topic has for its subtopics (a) personal aspects, i.e., activities, etc., done alone, (b) social aspects, i.e., activities with others, (c) physical aspects, i.e., sports, hobbies, etc. Items representative of this category are

³ Gratitude is expressed to Harrison G. Gough for his direction and assistance in developing the items of the VLP *Q* Sort. His insistence upon refinement of items and suggestions in working out the final wording of the items were major factors in the production of the VLP *Q* deck.

"Sociable, likes to be with people, to attend parties and gatherings of friends and acquaintances" and "Likes hobbies or leisure time activities with a technical aspect, e.g., developing and printing film, working with hi-fi equipment, etc."

2. The occupational topic divides into (a) socio-economic concerns, (b) satisfactions, (c) personal orientation toward work, (d) mobility-versatility, (e) talents and skills, (f) implications of vocational adjustment. This topic is covered by items such as "Financial and occupational security is of great importance" and "Prefers problems which require precise, exact, and logical thinking."

3. The personality characteristics topic involves (a) adjustment, (b) sense of responsibility, (c) self-evaluation and self-improvement, (d) values, (e) role considerations. The items of this category try to get at the personal motivation and dynamics which give the individual an orientation toward life, e.g., "Takes duties and obligations seriously, accepts responsibility for self" and "Aesthetic and cultural values are crucial to the development of character and personal maturity."

4. The interpersonal topic has as its subtopics (a) family and home concerns, (b) relationship to neighbor, (c) concerns about society, and (d) relationships with clients, etc. These ideas are exemplified in items such as "The true stature of man is indicated by his interest in and concern for the welfare and happiness of others" and "In vocational planning was always able to count on the support and encouragement of family."

The 70 items are distributed in the deck as follows: (a) recreational category, 12 items; (b) occupational category, 19 items; (c) personality characteristics category, 23 items; and (d) interpersonal category, 16 items.⁴

Studies of the VLP *Q*-sort deck have shown that all but two of the items tend to have a mean placement in the neutral or middle category with a distribution standard deviation of three–five categories. Test-retest reliability studies have provided reliability coefficients of .74 for females and .77 for males over a 2-wk. interval and of .73 for both males and females over a 4-wk. interval. Cluster analyses of the deck based on the groups used in this investigation found one cluster of two items for each of the individual groups and one cluster of four items when all 343 of the possible Ss for this investigation were used. Hence 66–68 items operate with considerable independence of one another in the deck.

In the initial study significant differences in item placement were determined, following Strong's (1943) method on the SVIB, by testing each group against the combined other three groups which were designated as the "professional men in general" sample. In this manner 50 Ss in one group were compared with the 150 Ss in the other three groups. Differential item placement was determined by *t* test using the Institute of Human Development *t*-test program for

⁴ A full discussion of the development of the VLP *Q* Sort can be seen in Kunert, 1965.

the IBM 7090 computer at the University of California Computer Center. The initial sample was then analyzed in terms of the significant items found for each professional group.

For the cross-validation study, differential *Q* sorts (*DQs*) were developed from the initial study results. The *DQs* were formed by ranking the items in each vocational group in the descending order of the *t* values derived in the initial investigation. The 4 items with the largest positive *t* values were then assigned to the most characteristic category, the next 9 items in order to the quite characteristic category, etc., until all 70 items were assigned to the categories in accordance with the distribution used for the VLP *Q* sort. The *Q* sort so developed became the modal or *DQ* sort for that vocational group. The *Q*-sort distributions of the 100 cross-validation *Ss* (25 *Ss* from each of the four vocational groups) were then correlated with each of the four *DQs*. These correlations were treated as the individual's scores on the *DQs* and analyses were performed to determine whether the *DQs* would differentiate the appropriate *DQ* vocational group from the other three groups.

RESULTS

Initial Study

In this investigation each group was compared with the other three groups and the number of differentiating items was determined for each group by means of the *t*-test procedure. For a *Q*-sort deck of 70 items one would expect by chance alone that 1 item would be significant at the .01 level and 4 items would be significant at the .05 level.

The *t*-test procedure showed that for the subsample of law students 11 out of the 70 items were significant with a *p* value of .01 or less. Eight more items (19 in all, 27% of the items) were significant with a *p* value of .05 or less.⁵

⁵ Eleven pages comprising Tables A through D which give the significant items for each vocational

For the subsample of medical students 5 out of the 70 items were significant with a *p* value of .01 or less. Eight more items (13 in all, 18% of the items) were significant with a *p* value of .05 or less.

For the subsample of divinity and theology students 29 out of the 70 items were significant with a *p* value of .01 or less, while 6 more items (35 in all, 50% of the items) were significant with a *p* value of .05 or less.

Finally, for the subsample of engineering students 21 of the 70 items were significant at the .01 level of significance or less and 13 more items (34 in all, 49% of the items) were significant at the .05 level of significance or less.

In the case of each subsample it can be seen that the number of significantly differentiating items exceeds by a considerable degree the chance expectations of one significant item at the .01 level of significance and the chance expectation of four items significant at the .05 level of significance.

Cross-validation study

In the cross-validation study the correlation between the *Ss Q* sort and the *DQ* was taken as his score on the *DQ*. Each *S* obtained a score on each of the four *DQs*. It was hypothesized that the law students would

group have been deposited with the National Auxiliary Publication Service. Order Document 00195 from National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 22 West 34th Street, New York, New York 10001. Remit in advance \$3.00 for photocopies or \$1.00 for microfiche and make checks payable to: Research and Microfilm Publication, Inc.

TABLE 1
t-TEST VALUES COMPARING PROFESSIONAL GROUPS TWO AT A TIME ON *DQ* SCALES

| Group | On scale for | Group | | | |
|----------------|--------------|---------|---------|----------|----------|
| | | 1 | 2 | 3 | 4 |
| 1. Law | Law | | 4.70*** | 13.24*** | 2.88** |
| 2. Medical | Medical | 5.69*** | | 7.97*** | 1.80* |
| 3. Theological | Theological | 9.40*** | 7.53*** | | 13.07*** |
| 4. Engineering | Engineering | 3.85*** | 3.75*** | 10.34*** | |

* *p* = .05 (*t* = 1.71 with 24 *df*).
** *p* = .01 (*t* = 2.49 with 24 *df*).
*** *p* = .001 (*t* = 3.74 with 24 *df*).

TABLE 2
CONTINGENCY TABLES FOR PROFESSIONAL GROUPS ON *DQ* SCALES

| Group | Law <i>DQ</i> Scale | | | Medical <i>DQ</i> Scale | | |
|--------|---------------------|-------|-------|-------------------------|-------|-------|
| | Law | M T E | Total | Medical | L T E | Total |
| Accept | 19 | 15 | 34 | 17 | 16 | 33 |
| Reject | 6 | 60 | 66 | 8 | 59 | 67 |
| Total | 25 | 75 | 100 | 25 | 75 | 100 |

| Group | Theological <i>DQ</i> Scale | | | Engineering <i>DQ</i> Scale | | |
|--------|-----------------------------|-------|-------|-----------------------------|-------|-------|
| | Theological | L M E | Total | Engineering | L M T | Total |
| Accept | 23 | 8 | 31 | 22 | 24 | 46 |
| Reject | 2 | 67 | 69 | 3 | 51 | 54 |
| Total | 25 | 75 | 100 | 25 | 75 | 100 |

have the highest scores on the law *DQ*, the medical students on the medical *DQ*, etc.

A *t*-test analysis of the scores of the four groups taken two at a time on each of the *DQ* scales gave the results found in Table 1.

In ten of the comparisons the *DQ* scales distinguished the scale appropriate group at the .001 level of significance. The law scale distinguished the law group from the engineering group at the .005 level of significance. The medical scale distinguished the medical group from the engineering group at the .05 level of significance.

The *DQ* scales also differentiated individuals according to their field of study. Analyses of these differences were based on frequency distributions of the scores and selected cutoff points on each of the *DQ* scales. The distribution of scores was analyzed by use of chi square. The contingency table data for this

analysis are given in Table 2. The results of the analyses are presented in Table 3.

The pattern of scores for each group and the differences among the groups are shown in Figure 1. All scores were transformed to *T* scores with a mean of 50 and a *SD* of 10 in order to make the scales comparable. As is evident from Figure 1, not only the scale scores but also the pattern of scores are useful

TABLE 3

CHI-SQUARE VALUES OBSERVED BETWEEN PROFESSIONAL GROUPS ON *DQ* SCALES

| Group | Scale | Group | | | |
|-----------------|-------|--------|--------|--------|--------|
| | | MTE | LTE | LME | LMT |
| Law (L) | DQL | 26.20* | | | |
| Medical (M) | DQM | | 18.47* | | |
| Theological (T) | DQT | | | 57.99* | |
| Engineering (E) | DQE | | | | 23.67* |

* $p < .001$ ($\chi^2 = 10.827$ with 1 *df*).

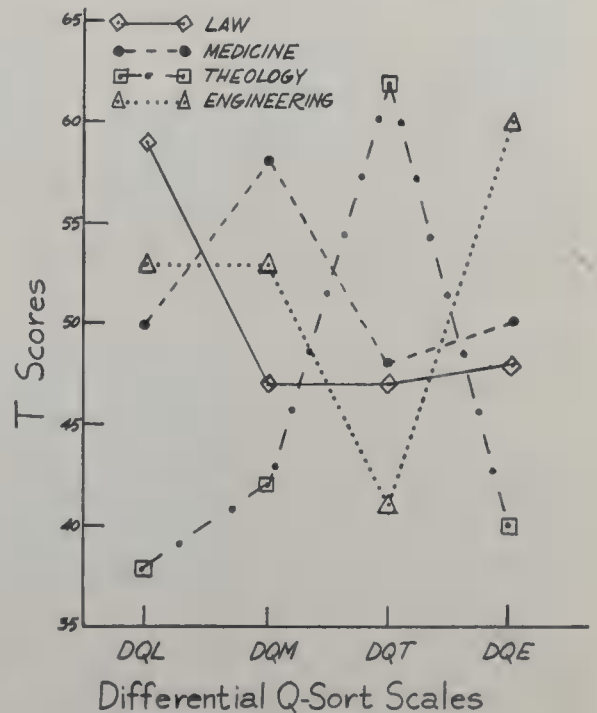


FIG. 1. Mean score patterns for four student groups on *DQ* scales.

in determining one's personality propensity for a particular vocational field.

DISCUSSION

The initial and cross-validation studies confirm the ability of the VLP *Q* Sort to discriminate students in the fields investigated. The initial study found a significant number of discriminating items for each group and ordered all of the items in a hierarchical progression on the basis of the *T* scores derived for each item. The cross-validation results showed that the *DQ* scales developed from the *T* scores of the initial study were able to discriminate both the groups and the individuals within the groups from one another. Thus the hypothesis of the investigation was confirmed.

An added advantage of the *Q*-sort method is that it showed the hierarchy of values within the groups as they gave differentiating emphases to their particular aims, values, and goals. The method also allowed for a wide range of individual variability as the *DQ* for each group gave the salient features of the group's personality profile and the individual sorts showed how this pattern was integrated idiographically within each *S*'s personality pattern.

Personality descriptions derived from the discriminating items of each group are presented below. The descriptions, however, are relative for they describe each group as it appears in comparison with the other three groups. To balance out the descriptions, differences between the relative and absolute *Q* sorts for each group will also be presented.

Law Group

The law student is concerned with himself, his position, and his status. He shows considerable interest and involvement in political affairs. He sees himself as perceptive and able to evaluate people, problems, and issues. He appears, however, more concerned with the implications of problems than with the people involved in them. He prefers to work with the intellectual aspects of problems rather than with the aspects which require precise and logical thinking. It also appears that duties and obligations are of secondary concern for

him. Religion also seems to hold a minimal position in his pattern. Relative to the other groups, he finds less meaning and fulfillment in his life and in his work. In fine his interests are largely intellectual and aesthetic in nature.

When the law group *Q* sort is considered by itself, duties and obligations are seen as important and life is meaningful for him. He accepts personal responsibility and is more reserved about his perceptive and evaluative abilities. He is also less active in expressing publicly his beliefs and ideas. Position and status are also less emphasized. He is, nevertheless, confident of his ability to cope with life's problems.

Medical Group

Family is of much importance for the medical student both in offering support and in providing a release from the stress and strain of life. He has little interest in being an organizer or leader of others, in getting involved in political affairs, and in impressing others. Indeed, he appears to judge others on the basis of surface traits and to relate to them in a rather reserved manner. He would appear to be quite controlled emotionally.

He is primarily involved in his work, and although he does not like to be overly restricted, yet he tends toward conservatism and is reluctant to try out new ideas, techniques, etc., in his work. He relies on logic and reason for making his decisions. Physical soundness is important to him and he enjoys competitive physical activities. In general, he is content with himself, his work, and his relationships with others.

When the medical *Q* sort is taken by itself, this student is seen to find life meaningful and to take his duties and obligations seriously. The importance of family and of hard work and constructive achievement are given less emphasis. He also sees himself as being less conservative and as being freer in relating to others.

Theological Group

The divinity or theology student places great importance on principles and beliefs both for guiding his daily life and for evaluating others. He is other-directed and deeply

concerned with helping others. He desires to correct social conditions and to lead and inspire others. He sees himself as being able to have anyone as a friend, and he works to increase the feeling of well-being both in others and in the world. Even his introspectiveness is used to understand others better, and he develops his skills, social and other, to be more effective in working with and interacting with others.

Yet his principles tend to lead him to be over-idealistic and even rigid in evaluating the world about him. He does not like all the tasks he may be assigned and is less intellectually involved in those which are less appealing. He finds his primary rewards in his work and holds family, home, personal ambition, and prestige to be of secondary importance.

His intuitive qualities and his contact with his emotions, while generally helpful to him, also at times tend to interfere with his work, especially when he is faced with difficulty and frustration.

When he is seen by himself, the theology student appears a little more flexible, more alert to prestige and financial concerns, and better able to deal with difficulties and frustrations. Also he tends to esteem personal freedom more and to put less emphasis on the subjective and intuitive approach to life. Even he is not particularly fond of participating in church clubs, suppers, and fund drives. In general, life is very meaningful for him, and he has a strong sense of responsibility for himself, his duties, and his obligations.

Engineering Group

The engineering student sees himself as being intellectually quick, exact, orderly, and analytical. He can also integrate material well and then present it clearly and effectively to others. He thrives on challenging ideas and problems and competitively seeks to excel in his work. For him, hard work and constructive achievement provide the means to attain financial and occupational security and thus to offer his family the best in life. In all of these areas he shows an ingrained sense of responsibility.

He prefers hobbies requiring technical and mechanical skills. He has a preference for short stories and magazine articles rather than for longer books and serious nonfiction as his desire for quickness again seems to predominate.

He is sociable and enjoys being with others, but he does not wish to get involved in others' problems, nor in social or political problems. He is basically pragmatic with little interest in personal introspection or with abstract and aesthetic considerations. Religious beliefs and church activities occupy but a small part of his object-oriented world. His philosophy is one of "live and let live," and thus he avoids involvement in the lives and affairs of others and in turn expects the same from them in his own regard.

The picture varies little when the Engineering Group's *Q* sort is considered by itself. It shows less emphasis for hobbies requiring technical and mechanical skills and less emphasis on the concern for financial and occupational security. His interest in short stories, etc., is more moderate as is his picture of himself as evaluating problems quickly. Social concerns tend now to fall more in the neutral region than in the uncharacteristic region of his life pattern.

REMARKS

This investigation indicates that ipsative measures can discriminate the pattern of proactive personality dynamics operating within different vocational groups. Yet, while it has provided some thought-provoking results, its major value would seem to lie in the fact that it has shown that it is possible to tap the rational and motivational aspects of people's lives. The *Q*-sort method joined with a *Q* deck developed according to one's major theoretical positions is capable of bringing together the rational and empirical approaches to psychology to provide data of both practical and theoretical import. With this method it would appear that difficult areas of personality theory can be put to the test. It might even make possible the attainment of a more or less all-embracing theory of personality upon which a stable, scientific structure can be built.

REFERENCES

- DARLEY, J. G., & HAGENAH, T. *Vocational interest measurement*. Minneapolis: University of Minnesota Press, 1955.
- ENGLANDER, M. E. A psychological analysis of vocational choice: Teaching. *Journal of Counseling Psychology*, 1960, 7, 257-264.
- ENGLANDER, M. E. Q sort: A means to explore vocational choice. *Educational and Psychological Measurement*, 1961, 21, 597-605.
- GINZBERG, E., GINSBURG, S. W., AXELRAD, S., & HERMA, J. L. *Occupational choice*. New York: Columbia University Press, 1951.
- KUNERT, K. M. The psychological concomitants and determinants of vocational choice. (Doctoral dissertation, University of California, Berkeley) Ann Arbor, Mich.: University Microfilms, 1965. No. 65-8198.
- MORRISON, R. L. Self-concept implementation in occupational choices. *Journal of Counseling Psychology*, 1962, 9, 255-260.
- NEFF, W. S., & HELFAND, A. A Q-sort instrument to assess the meaning of work. *Journal of Counseling Psychology*, 1963, 10, 139-145.
- ROE, A. *The psychology of occupations*. New York: Wiley, 1956.
- STRONG, E. K., JR. *Vocational interests of men and women*. Stanford: Stanford University Press, 1943.
- SUPER, D. E. A theory of vocational development. *American Psychologist*, 1953, 8, 185-190.
- SUPER, D. E. *The psychology of careers*. New York: Harper, 1957.

(Received April 2, 1968)

EFFECT OF PERCEIVED EXPERTNESS UPON CREATIVITY OF MEMBERS OF BRAINSTORMING GROUPS

PANAYIOTA A. COLLAROS AND LYNN R. ANDERSON¹

Wayne State University

Perceived expertise of other members may make a brainstorming group less effective than the pooled results of members working alone. In ■■ All Experts condition, each member of a brainstorming group was told that all other members had previously worked in such groups; while in a One Expert condition it was stated that only one member (unidentified) had this experience. No information was given in a Control condition. The Ss felt more inhibited in the All Experts than in the One Expert condition which, in turn, had more inhibition than the control. Originality and practicality of ideas varied according to the degree of felt inhibition, the Control condition had the highest originality and practicality scores, followed by the One Expert condition and then the All Experts condition.

Osborn devised the brainstorming procedure to create a free and uninhibited atmosphere which would increase the creativity of group members. The main features of the brainstorming procedure described by Osborn (1957) are judicial judgment of ideas is ruled out; freewheeling ideas are welcomed; quantity is wanted; and combinations and improvement of ideas are sought. Osborn's own research, for example, showed that engineers were able to produce "44% more worthwhile ideas" using a group brainstorming technique than when the members worked alone using other than brainstorming techniques (Osborn, 1957). Osborn concluded that group participation under the brainstorming conditions can improve significantly the creativity of the group members.

One of the essential features of the brainstorming technique is that novel, offbeat ideas produced by one member of the group can suggest even more novel or original ideas to another member of the group. This effect should be especially evident if the members can contribute the unique ideas in a nonevaluative atmosphere without fear of censure from others. It is this mutual stimulation in a nonevaluative atmosphere that could, purportedly, give a qualitatively different dimension to the group ideas compared to the pooled ideas of the individuals working alone.

However, recent research by Taylor, Berry, and Block (1958) suggests that group brainstorming actually inhibits the creativity of the individual. The study was replicated, successfully, by Dunnette, Campbell, and Jaastad, in 1963.²

One plausible reason which might partially account for the failure of group brainstorming to prove superior to pooled individual brainstorming is the "self-weighting" effect discussed by Kelley and Thibaut (1954). According to these authors, members of groups should participate and evaluate their own ideas according to their felt competency within the group. Even in the brainstorming group where such effects should be minimized, the individual probably will participate to the extent that he feels as capable as other members and to the extent that he is familiar with the brainstorming technique. Although

¹ Requests for reprints should be sent to Lynn R. Anderson, Department of Psychology, Wayne State University, Detroit, Michigan 48202.

² Although studies by Taylor et al. (1958) and Dunnette et al. (1963) found that individuals brainstorming in a group were less creative than individuals brainstorming alone, the studies do not present an accurate test of the brainstorming procedure since Osborn devised the brainstorming procedure to increase the creativity of groups (groups in Osborn's advertising agency, to be exact). Neither study compared the efficacy of group brainstorming with conventional techniques of group problemsolving. Research which has compared group brainstorming to previous, conventional methods of group problemsolving, has usually found that, indeed, more ideas and more creative ideas are produced in the brainstorming groups (Meadow, Parnes, and Reese, 1959; Weisskopf-Joelson and Eliseo, 1961).

the brainstorming instructions direct the members to make no evaluative judgments of any ideas which are offered, it probably is impossible to eliminate covert judgments which are made but are not expressed as overt criticisms. Consequently, status or competency differences among the members probably create inhibitions which, in turn, produce a decrease in the number of ideas produced by the low status or less competent members. This inhibition effect may, in part, account for the fact that the group brainstorming is usually found to be inferior to the statistical pooling of the individual member's creative efforts.

In the present study the perceived expertness of the other group members was varied in order to examine the effects such perception would have upon the "self-weighting" of the members' contributions and, subsequently, upon the creativity of the brainstorming group.

METHOD

Subjects

The Ss were 240 undergraduate (120 males) students enrolled in classes of introductory psychology at a large midwestern university.

Instructions

Three inhibition conditions were created in the experiment through the manipulation of perceived expertness: no perceived experts in the group, one perceived expert, and all perceived experts. In each condition 80 Ss were assigned to 4-man groups. Assignments to the 20 brainstorming groups in each condition was random, except that all members of the group were of the same sex and were not acquainted with one another before the experiment.

Before each brainstorming session started, Ss met with *E* (the same *E* ran all Ss), where they were given the instruction sheets. These sheets included general brainstorming instructions as well as information designed to create the specific perceived expertness manipulations.

The instruction sheet for the groups in the All Perceived Experts condition included a paragraph which read

Read this carefully as you are the only person in this group who has not participated in a brainstorming session before. All other members are familiar with the procedure to be followed.

In this condition all four members were led to believe that the other three members had had some previous experience with the brainstorming procedure.

The instruction sheet for the One Perceived Ex-

pert condition included a similar paragraph which read

Only one person in each group has participated in a brainstorming session before. The rest should read the following very carefully to become familiar with the procedure to be followed.

Again all four members in the group received the same instruction sheets so that each one thought that there would be one member with brainstorming experience in his group.

The Ss in the Control or No Perceived Experts condition received instruction sheets which included no information about the other members of the group. Therefore, each *S* perceived the others as being equal to himself in brainstorming expertise.

The instruction sheets in all but the Control conditions were mimeographed on different colored sheets to enhance the fact that differences did exist among the members of the group. The instruction sheets were collected before Ss began working on the brainstorming problem so that members of each group could not compare instructions. The *E* also tried to emphasize the instruction and the expert manipulation with such oral comments as "The ones who are here for the first time, please read the instructions very carefully," or "If you have any questions, ask me now, because the members in your group who have experience with brainstorming have been instructed not to answer any questions during the session."

Procedure

After Ss had read the instruction sheets, they were assigned to their brainstorming group which met in an isolated room. The brainstorming problem was taken from a group creativity study by Triandis, Hall, and Ewen (1965). It read "How can a person of average ability achieve fame and immortality though he does not possess any particular talents?" The Ss recorded their answers on a sheet which also contained three 5-step scales which *S* could use to rate (at a later time) the "originality," "creativity," and "practicality" of each idea offered. The groups were given unlimited time to work on the problem so that the discussion was terminated only after the group had no more ideas to offer. After the session *E* collected the answer sheets and distributed a postmeeting questionnaire to each member. The questionnaire included items intended to measure *S*'s feelings about the brainstorming session and his participation in the session.

The measure of inhibition was taken from *S*'s rating of his own feeling about the amount of inhibition in his group and also by having *S* list any solutions to the problem which he did not contribute during the group discussion. The rationale behind this procedure is that since *S* had unlimited time during the group discussion to list his ideas, the ideas which he would list after the session may have been ideas he had held back during the group session. Hence one inhibition score was computed for each *S* based on a ratio of the number of

ideas offered after the session over the total number of ideas listed both during the session and after the brainstorming session. This score is based on a ratio score found in the study of group creativity by Triandis et al. (1965).

RESULTS

The effectiveness of the perceived expertness manipulation can be seen from the items on the postmeeting questionnaire which indicate Ss' own feelings about the group session. Four items were rated on a 5-point scale and were stated as: "Were you reluctant in offering an idea for fear of criticism from other members?" "Were you at all inhibited due to the presence of others who had more experience with brainstorming?" "When you offered an idea that was 'way out,' did you sense a certain disapproval from other members, although no overt criticism was expressed?" and "Did such fear of possible disapproval from other members make you withhold any ideas?" The analyses of variance which were conducted on these ratings (see Table 1) showed that the perception of inhibition in the All Experts condition was always significantly higher than the perceived inhibition in the Control condition; the One Expert condition had a moderate amount of perceived inhibition which fell between the two extreme conditions. In no instance is there a reversal from the pattern of inhibition which was predicted from the expertness manipulation. These data indicate that the

expertness of the other group members did make the individual member feel reluctant to contribute all the ideas which came to mind.

A direct test of the inhibition manipulation can also be obtained from a count of the number of ideas produced in each of the expertness conditions. The Ss in the No Experts condition produced an average of 16.06 ideas, which is significantly higher ($p < .01$) than the average of 9.25 ideas produced by Ss in the One Expert condition. In turn, the mean of the One Expert condition was significantly higher ($p < .01$) than the mean of the All Expert condition (4.71). The ratio inhibition score was computed from a count of the number of ideas produced alone in the postmeeting session compared to the total number of ideas which each *S* produced (alone plus group ideas). This ratio score showed again that the expertness factor inhibited the number of ideas which Ss were willing to contribute. The ratio inhibition scores are also shown on Table 1, where it can be seen that the All Experts condition was significantly higher (more inhibition) than the One Expert, but this later condition was not significantly higher than the Control condition when this particular inhibition score was used.

Since Ss had rated their own ideas on the dimensions of creativity, originality, and practicality, it was possible to compare the "quality" of ideas produced in each of the inhibition conditions. These scores are shown

TABLE 1
MEAN INHIBITION SCORES IN THE VARIOUS EXPERT CONDITIONS

| Postmeeting questions | Expert condition | | |
|--|---------------------------------|--------------------------------|-------------------------------|
| | All Experts (<i>n</i> = 80) | One Expert (<i>n</i> = 80) | No Expert (<i>n</i> = 80) |
| "Reluctant" question | 2.72 ^a (.914) | 2.55 ^a (.674) | 1.36 (.580) |
| "Felt inhibited" question | 2.41 (.837) | 2.39 ^a (.741) | 1.70 (.879) |
| "Sensed disapproval" question | 3.07 (.938) | 2.45 ^a (.656) | 1.70 (.819) |
| "Withheld ideas" question | 2.97 ^a (.900) | 2.36 ^a (.663) | 1.62 (.722) |
| Ratio score (No. ideas alone/No. ideas alone + ideas in group) | .364 (.256) | .057 (.009) | .019 (.004) |

Note.—The numbers in parentheses are the standard deviations for the variables.

^a Indicates the mean is significantly higher ($p < .01$) than the mean in the column to the right.

TABLE 2

RATINGS OF THE IDEAS PRODUCED IN EACH INHIBITION CONDITION

| | Expert condition | | |
|--------------|--------------------------------|--------------------------------|---------------------------------|
| | No Experts (<i>n</i> = 80) | One Expert (<i>n</i> = 80) | All Experts (<i>n</i> = 80) |
| Creativity | 2.97 | 3.06 ^a | 2.69 |
| Originality | 3.64 ^a | 3.39 ^a | 3.11 |
| Practicality | 3.36 ^a | 3.25 ^a | 3.08 |

^a Indicates the mean is significantly higher ($p < .05$) than the mean in the column to the right.

in Table 2 and indicate quite clearly the detrimental effect of the inhibition manipulation. The Ss in the Control condition rated their ideas higher on practicality and originality than did Ss in the All Experts condition. The scales were highly intercorrelated and since no specific definitions of these items were given, the ratings probably represent a gross evaluative rating.³

One final item on the postmeeting questionnaire provides some additional data regarding the effects of the expert manipulation. This item reads "As a personal experience, did you find the session pleasant?" In the All Experts condition the mean rating was 2.17 (on a 5-point scale) which is significantly lower ($p < .01$) than the mean rating of the One Expert condition (2.59) which was, in turn, significantly lower ($p < .01$) than the mean pleasantness rating in the Control condition (3.81).

DISCUSSION

The major problem of this study was the effect of perceived expertness of others upon individual creativity in brainstorming groups. The results showed quite conclusively that individuals were reluctant to contribute all of their ideas when they were in groups with members who were thought to have had previous training and experience was the brainstorming procedure. Since the creativity of the brainstorming group is highly depend-

ent upon all members in the group contributing all possible ideas and solutions to the problem, total group creativity will be noticeably lessened when one member of the group appears to be more expert than the other members. The members of the experimental groups specifically stated that they "felt inhibited" during the brainstorming session, and they also evidenced this fact by listing in the postsession "alone" situation many ideas which they had not contributed to the group discussion.

The inhibition which the members of the "expert" groups felt may be due to the implied threat of the more knowledgeable members (Hoffman, 1965) and the possibility of censure by these expert members. Although the brainstorming instructions specified that no criticism was allowed, Ss in the expert group did note in the postmeeting questionnaire (see Table 1) that they "sensed disapproval from other members." Consequently these group members did not contribute a large portion of ideas which came to mind; thus creativity in the inhibition conditions was significantly lower than in conditions where members did not feel the threat of censure by more knowledgeable members.

The results of the study confirm the findings of other investigators regarding the effects of status differences and ability differences on group problemsolving (Hoffman, 1965). For example, Torrance (1955) has found that the low status person may be inhibited and "go along" with opinions expressed by the high status person even though he feels his own opinions are better. Mausner (1954) also found that Ss conform to their partner's erroneous judgments more often when the partner has had previously successful experience with the task than when the partner had had previously unsuccessful experience. Our results indicate that, indeed, the presence of an expert in the brainstorming group made the individual feel inhibited and subsequently caused him to contribute only a few of his ideas to the group problem.

The individuals who were brainstorming in groups with purported experts also rated their groups as being less pleasant than groups without expert members. These ratings of unpleasantness made by Ss in the "expert"

³ It should also be noted that Ss' ideas were also rated by two outside judges using the same scales. These ratings were highly correlated with Ss' own ratings ($p < .001$) and, consequently, Ss' own ratings were used as the criterion scores.

conditions and the fact that at the end of the brainstorming session these members still had many ideas which were unexpressed undoubtedly reflect Ss' reluctance to contribute all of their ideas. Members of the Control condition expressed a much larger proportion of their ideas and, subsequently, they felt that the group session was much more pleasant than did Ss in the "expert" conditions.

Overall the results of the study indicate that social factors inherent in unequal status structures within the group are detrimental to member creativity even though brainstorming instructions are given. Group members feel threatened and inhibited by the presence of more knowledgeable members, consequently, the less expert members contribute few of their ideas and suggestions, that is, group members "self-weight" their own contributions according to their previous experience with the brainstorming technique (Kelley and Thibaut, 1954). This finding may explain, in part, why many studies which have compared the creativity of individuals who are brainstorming alone and in groups unanimously endorse the superiority of the "alone" condition.

The implications of such a finding would suggest that careful attention should be given to the selection of members for the brainstorming group by making previous experience as similar as possible. When research involves the intact group consisting of members with extremely different backgrounds in brainstorming techniques, then effort should be made to conceal or minimize these differences within the group.

REFERENCES

- DUNNETTE, M. D., CAMPBELL, J., & JAASTAD, E. The effect of group brainstorming effectiveness for two industrial samples. *Journal of Applied Psychology*, 1963, 47, 30-37.
- HOFFMAN, L. R. Group problem solving. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. II. New York: Academic Press, 1965.
- HOROWITZ, M. The recall of interrupted group tasks: An experimental study of individual motivation in relation to group goals. In D. Cartwright and A. Zander (Eds.), *Group dynamics*. Evanston, Ill.: Row, Peterson and Co., 1960.
- KELLEY, H. H., & THIBAUT, J. W. Experimental studies of group problem solving and process. In G. Lindzey (Ed.), *Handbook of social psychology*. Reading, Mass.: Addison-Wesley, 1954.
- MAUSNER, B. The effect of one partner's success in a relevant task on the interaction of observer pairs. *Journal of Abnormal and Social Psychology*, 1954, 49, 557-560.
- MEADOW, A., PARNES, S. J., & REESE, M. Influence of brainstorming instructions and problem sequence on a creative problem solving test. *Journal of Applied Psychology*, 1959, 43, 413-416.
- OSBORN, A. F. *Applied imagination*. New York: Scribner, 1957.
- TAYLOR, D. W., BERRY, P. C., & BLOCK, C. H. Does group participation when using brainstorming facilitate or inhibit creative thinking? *Administrative Science Quarterly*, 1958, 3, 23-47.
- TRIANDIS, H. C., HALL, E. R., & EWEN, R. B. Member heterogeneity and dyadic creativity. *Human Relations*, 1965, 18, 33-55.
- TORRANCE, E. P. Some consequences of power differences on decision making in permanent and temporary three-man groups. In A. P. Hare, E. F. Borgatta, and R. F. Bales (Eds.) *Small groups: studies in social interaction*. New York: Knopf, 1955.
- WEISSKOPF-JOELSON, E., & ELISEO, T. S. An experimental study of the effectiveness of brainstorming. *Journal of Applied Psychology*, 1961, 45, 45-59.

(Received April 4, 1968)

VALIDITY OF ESTIMATES BY CLERICAL PERSONNEL OF JOB TIME PROPORTIONS

STEPHEN J. CARROLL, JR.,¹ AND WILLIAM H. TAYLOR, JR.²

Department of Business Administration, University of Maryland

This study compares the estimated time allocations of 16 clerical workers with their actual time allocations as determined by work sampling procedures carried out surreptitiously over a 2-wk. period. The average correlation between estimated and actual time allocations for individual workers was .88. Only 2 of the 16 correlations were under .80. The biggest difference between an estimated and actual time allocation for a particular work activity was 6% for the group of workers as a whole. This study seems to indicate that time estimates from rank and file workers can be accurate enough to be useful in employee recruitment, selection, evaluation, training, and compensation.

Only a few studies have been completed which have focused on the accuracy of employee estimates of how they allocate their work time among various work activities. Since time estimates are used in job analysis and are certainly the cheapest and simplest of the methods used to measure work, the validity of the information obtained by this method should be of interest to all those involved in the derivation and use of job information.

In a study by Stogdill and Shartle (1955), estimates of the time spent in various work activities made by 34 naval officers were compared to a log of time spent in work activities maintained by the officers for three days. It was found there was a fairly high relationship between estimated and actual logged time for specific work activities such as talking, reading, writing reports, and operating machines. More subjective activities such as planning and reflection were less accurately estimated. Mahoney, Jerdee, and Carroll (1963) in a study of 4 managers in one company and 28 managers in another found that job classifications based on time estimates correlated moderately well with job classifications based on time allocation as determined by work sampling. Finally a study of 232 technical men by Hinrichs (1964) indicated that estimates of the proportion of

time spent in communication activities were very close (within 5%) to the time proportions obtained with work sampling.

The purpose of the present study was to evaluate the accuracy of time estimates made by a group of lower level workers. Previous studies have focused only on higher level workers, and the extent to which their findings may be generalized is still virtually unknown.

METHOD

At the start of the study, 16 clerical workers were asked to estimate the proportion of time each spent on various job activities during a routine workday. Next an independent observer obtained random observations of their work activities 16 times a day for a period of 2 wk. These random observations were then classified by the observer into appropriate work activity categories. The proportion of observations falling in each category then formed the basis for inferences about the actual time allocations as determined by work sampling. The participants were unaware that their work activities were being observed since they were not told and since the observer regularly spent a considerable amount of time in the office. While this procedure violates normal work sampling procedure it was considered justified in this case. No information detrimental to any individual was given to management.

RESULTS

Table 1 presents the results of this study. Although the differences between the estimated time allocations and the time allocations determined by work sampling are significantly different in a statistical sense (chi-square, .01 level), the actual differences are not large, being at the most 6% (for machine operation). Idle and personal time which

¹ Requests for reprints should be sent to Stephen J. Carroll, Jr., Department of Business Administration, University of Maryland, College Park, Maryland 20740.

² Now with the United States Naval Weapons Station, Charleston, South Carolina.

TABLE 1

PROPORTION OF TIME SPENT IN VARIOUS WORK ACTIVITIES AS DETERMINED
BY ESTIMATES AND BY WORK SAMPLING

| Work activity | Job time proportions as determined by work sampling | | Job time propor- tions as determined by estimates | Differences in average time allocation |
|---|---|-------|---|--|
| | <i>n</i> | % | % | % |
| Conversation | 246 | 10.3 | 7.5 | 2.8 |
| Filing | 202 | 8.5 | 10.8 | 2.3 |
| Idle and Personal | 169 | 7.1 | 2.4 | 4.7 |
| Machine operation | 155 | 6.5 | 12.4 | 5.9 |
| Mail handling | 11 | .5 | .8 | .3 |
| Telephone | 108 | 4.6 | 7.9 | 3.3 |
| Typing | 179 | 7.5 | 6.3 | 1.2 |
| Walking | 89 | 3.7 | 5.0 | 1.3 |
| Writing, research and review ^a | 1074 | 45.1 | 46.8 | 1.7 |
| Other | 75 | 3.2 | 0.0 | 3.2 |
| Unknown | 72 | 3.0 | 0.0 | 3.0 |
| Totals | 2380 | 100.0 | 100.0 | |

^a These were originally separated in the study but were combined because the observer and Ss had difficulty in differentiating among them.

might have been expected to be the least accurately estimated activity since employees would be most sensitive about this was off only 5%. Machine operation might have been off the most because this was the activity to

which the idle and personal underestimation was shifted. The proportion of time taken up by these two activities together is about 14% when determined by both estimates and work sampling. Attention should be directed also at the "other" and "unknown" work activity categories, for these categories obviously could not be used at all in the time estimates very well. This would mean that the time allocations for the estimates would have to differ from the time allocations from work sampling by at least the proportion of observations classified into these two categories. Keeping this in mind, it appears the two time allocations as determined by two different methods are quite similar.

This similarity in time allocations between estimated and actual is also true for individuals as well as for the group as a whole. This is indicated in Table 2 which presents the correlation coefficients between estimated and actual job activity times for each of the 16 employees studied. Only two of these correlation coefficients are below .80 and only one is not significant at the .01 level.

DISCUSSION

The study indicated that time estimates made by personnel on a lower level clerical

TABLE 2

RELATIONSHIP BETWEEN TIME ALLOCATIONS AS
DETERMINED BY ESTIMATES AND BY WORK
SAMPLING FOR 16 SUBJECTS

| Employee | Correlation ^a (Pearsonian) |
|----------|--|
| A | .80 |
| B | .98 |
| C | .92 |
| D | .99 |
| E | .99 |
| F | .99 |
| G | .82 |
| H | .97 |
| I | .98 |
| J | .99 |
| K | .98 |
| L | .99 |
| M | .80 |
| N | .22 |
| O | .75 |
| P | .90 |
| Average | .88 |

job were at least as accurate as those reported for personnel on several higher level jobs. The study also would seem to indicate that time estimates from a group of such personnel can be of value since they are so easy to obtain and can be accurate enough to serve as a general guide to the nature of the work performed on various jobs. For example, such job information would certainly prove useful in employee recruitment, selection, evaluation, training, and compensation.

REFERENCES

- HINRICHS, J. R. Communications activity of industrial research personnel. *Personnel Psychology*, 1964, 17(2), 193-204.
- MAHONEY, T. A., JERDEE, T. H., AND CARROLL, S. J., JR. Development of Managerial Performance. A Research Approach. *Monograph C-9*, Southwestern, Cincinnati, 1963.
- STODDILL, R. M., AND SHARTLE, C. L. Methods in the Study of Administrative Leadership. *Research Monograph 80*, Bureau of Business Research, Ohio State University, Columbus, 1955.

(Received April 10, 1968)

journal of applied psychology

Table of Contents

| | |
|--|-----|
| Task Structure, Work Structure, and Team Performance. <i>James C. Naylor and Terry L. Dickinson</i> | 167 |
| Motivation of Research and Development Entrepreneurs: Determinants of Company Success. <i>Herbert A. Wainer and Irwin M. Rubin</i> | 178 |
| Ratee Relevance in Peer Nominations. <i>Frank T. Passini and Warren T. Norman</i> | 185 |
| Self-Esteem as a Moderator in Vocational Choice: Replications and Extensions. <i>Abraham K. Korman</i> | 188 |
| Predictive Power over Ten Years of Measured Social Service and Scientific Interests among College Women. <i>Lenore W. Harmon</i> | 193 |
| Cognitive, Noncognitive, and Environmental Correlates of Mechanical Ingenuity. <i>W. A. Owens</i> | 199 |
| Architecture School Performance Predicted from ASAT, Intellectual, and Nonintellectual Measures. <i>Clifford E. Lunneborg and Patricia W. Lunneborg</i> | 209 |
| Relationship between Interview Structure and Interinterviewer Reliability in an Employment Situation. <i>Donald P. Schwab and Herbert G. Heneman III</i> | 214 |
| Automated Instruction for Vigilance Training. <i>Dennis A. Attwood and Earl L. Wiener</i> | 218 |
| Knowledge of Performance as an Incentive in Repetitive Industrial Work. <i>P. S. Hundal</i> | 224 |
| Work Shift, Occupational Status, and the Perception of Job Prestige. <i>Ronald H. Bohr and Arnold B. Swertloff</i> | 227 |
| Comparative Study of Need Satisfactions in Governmental and Business Hierarchies. <i>J. B. Rhinehart, R. P. Barrell, A. S. DeWolfe, J. E. Griffin, and F. E. Spaner</i> | 230 |
| Beyond Parkinson's Law: III. Effect of Protractive and Contractive Distractions on the Wasting of Time on Subsequent Tasks. <i>David Landy, Kathleen McCue, and Elliott Aronson</i> | 236 |
| Examination of Some Measures of Creative Ability by the Multitrait-Multimethod Matrix. <i>P. Goodman, J. Furcon, and J. Rose</i> | 240 |
| Assymetrical Transfer in Reading Texts Produced by Teleprinter and by Typewriter. <i>E. C. Poulton</i> | 244 |
| Predicting Vocational Success for Neuropsychiatric Patients with the Edwards Personal Preference Schedule. <i>Allen Goss</i> | 250 |
| Relation between Embedded Figures Test Performance and Simulator Behavior. <i>Gerald V. Barrett, Carl L. Thornton, and Patrick A. Cabe</i> | 253 |
| List of Manuscripts Accepted. | 192 |

KENNETH E. CLARK
University of Rochester

Consulting Editors

GEORGE E. BRIGGS
Ohio State University

MARVIN D. DUNNETTE
University of Minnesota

NORMAN FREDERIKSEN
Educational Testing Service

LEONARD D. GOODSTEIN
University of Cincinnati

EDWIN R. HENRY
University of Rochester

JOHN HOLLAND
American College Testing Program

CLIFFORD E. JURGENSEN
Minneapolis Gas Company

DONALD G. MARQUIS
Massachusetts Institute of Technology

LAURENCE S. MCGAUGHAN
University of Houston

QUINN MCNEMAR
University of Texas

WARREN T. NORMAN
University of Michigan

HAROLD F. ROTHE
Belott Corporation

THOMAS A. RYAN
Cornell University

ALEXANDER G. WESMAN
Psychological Corporation

CLARE L. WILSON
Instituto Centroamericano de Administracion de Empresas

This journal gives primary consideration to original investigations in any field of applied psychology except clinical psychology, although a descriptive or theoretical article may be accepted if it represents a special contribution in an applied field. Quantitative investigations of interest or value to psychologists working in the following broad fields will be considered: personnel research in business, industry, and government; engineering psychology; industrial working conditions; research on opinion and morale factors; job analysis and classification research; market and advertising research; vocational and educational prognosis, diagnosis, and guidance at the secondary and college level.

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or they may be obtained from the Editor or from the APA Central Office.

Manuscripts should be addressed to the Editor:

Dr. Kenneth E. Clark, Dean
College of Arts and Science
302 Morey Hall
University of Rochester
Rochester, N. Y. 14627

Authors are strongly advised to follow the general directions given in the *Publication Manual of the American Psychological Association* (1967 Revision). Special attention should be given to the section on the preparation of the references (p. 39). In reference lists, give journal titles in full; do not abbreviate. *All copy must be double spaced, including the references.* All manuscripts should be submitted in *duplicate*, one of which should be an original typed copy; author's name should appear only on title page. Dittoed and mimeographed copies are not acceptable and will not be considered. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail and to check carefully the typing of the final copy.

The following policies govern reprinting or other forms of reproduction of APA-copyrighted journal materials. Permission by the American Psychological Association must be requested for reprinting all tables, figures, and text of more than 500 words (total from one article) in length. There will be a charge of \$5.00 per page for such permission except when the requester is the author of the article to be reprinted, or the article is to be reproduced in a limited number of copies solely for instructional purposes.

Published bimonthly by the
American Psychological Association
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W.
Washington, D. C. 20036

\$10.00 per volume

\$2.00 per issue

HELEN ORR
Managing Editor

CLAIRE PILLIOD
Technical Editor

HENRY S. SALMON
Advertising and
Business Manager

ROBERT D. GORMAN
Subscription Manager

Subscriptions, orders, and business communications should be addressed to the American Psychological Association, 1200 Seventeenth St., N. W., Washington, D. C. 20036. Address changes must reach the subscription office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Second-class postage paid at Lancaster, Pa.

TASK STRUCTURE, WORK STRUCTURE, AND TEAM PERFORMANCE¹

JAMES C. NAYLOR AND TERRY L. DICKINSON

Ohio State University

Three different work structures were examined factorially with two levels of task structure and two levels of task organization using two-man teams in a multiple-cue inference task in an initial test of the Dickinson-Naylor taxonomy of team performance. All teams performed for 200 trials. Task structure significantly influenced team achievement, consistency, and matching, while task organization influenced only team achievement and matching behavior. Work structure failed to show any effect upon performance except in terms of the degree to which team responses could be predicted from individual member responses.

A model of the relevant dimensions of team performance has been suggested by Dickinson and Naylor (1966) which can be stated as follows: Team performance = f_0 (Task structure, Work structure, Communication structure).

In this model, task structure is broadly conceived as the demand characteristics of the task to be accomplished. Also, it is viewed as a "fixed" dimension in the sense that if the demand characteristics change, the task itself, by definition, is also changed. More specifically, task structure is formally defined to be a function of the individual and joint demand characteristics of the separate task components, namely, component complexity, component organization, and component redundancy. Stated formally, Task structure = f_1 (Component complexity, Component organization, Component redundancy).

Each of the task structure characteristics may in turn be defined. Thus the complexity of a task component is defined in terms of its

information-processing and/or memory-storage demand requirements. Component organization is defined in terms of similar demands imposed by the total task due to the interrelationships existing between or among task components. Finally, task redundancy refers to the degree of overlap existing among the demands imposed by the several individual task components.

Turning now to the characteristic of work structure, Dickinson and Naylor (1966) have defined this dimension of team performance as the manner in which the task components are distributed among team members. Work structure, then, may be viewed as a subtask work assignment problem in the study of team performance. It involves (a) the definition of the operations to be performed, (b) the sequence in which these operations must occur, and (c) the way in which interaction among team members must occur (Naylor & Briggs, 1965).

The third and final dimension of team performance, communication structure, is defined in terms of the communication interrelationships which exist between team members. It is exceedingly important to note that in the Dickinson-Naylor model the communication structure dimension differs from the work and task structure dimensions in one very im-

¹ This research is based on the master's thesis of the second author and was supported by Research Grant GB-4987 from the National Science Foundation awarded to the first author for research on choice behavior in multiple-cue situations.

² Requests for reprints should be sent to James C. Naylor, Department of Psychology, Purdue University, Lafayette, Indiana 47907.

portant way—communication structure is viewed as being a dependent variable rather than an independent variable; that is, given any particular combination of work and task structure, team members themselves will develop (within the limits of that particular unique situation) a particular communication structure. This point of view is similar to that proposed by Faucheux and MacKenzie (1966) where they contend that the communication structure developed by a team is only one of many which could be developed under the existing conditions. They conclude that communication structure is a dependent variable intervening between task and behavior.

The importance of the work and task structure variables with respect to communication structure is that they are viewed as placing limitations on the type of communication structure which a team will develop. Thus the work structure may be such that several individuals are required to perform the same subtask. In this situation the potential for interaction among team members is at a maximum and thus greatly facilitates the opportunity for communication. Similarly, task structure may also be an influence on communication structure. For example, as the complexity of the team task increases, the number of subtasks a team member can perform may be reduced, possibly leading to the need for a highly developed communication structure. The organization of the task may also be important. At one extreme there may be little or no interrelationship between the subtask components of any two members, thus restricting the necessity for communication. At the other extreme, a substantial interrelationship between subtask components may mean that intermember communication is absolutely necessary for subtask success, resulting in a great deal of pressure on members to develop patterns of communication.

It should also be pointed out that the above definition of communication structure involves only those communication systems or networks worked out by team members and does not include aspects of communication among members which are intrinsic to the nature and design of the task. An example of this distinction is a recent study by Williges, Johnston, and Briggs (1966). In one condition of that

study, team members could obtain information about their partner's activity via a radar display while in the other condition they saw only their own targets and interceptors. Under the present taxonomy, these two conditions would be viewed as different work structures, and the communication structure would be different for the two conditions only if different interpersonal communication patterns were found to emerge as a function of the work structure differences. Incidentally, this apparently happened in the Williges et al. (1966) study—at least in terms of frequency of communication among members. This would again support the notion of communication structure as a variable dependent upon work and task structure.

Clearly the adequacy of the Dickinson-Naylor model depends upon the extent to which it can account for the performance of teams. The crucial issues, of course, are the task and work structures imposed upon the team, since as previously explained communication structure is typically not manipulated per se but is dependent upon the first two dimensions. The purpose of this initial research was to examine team performance as a function of both the work structure and task structure variables.

METHOD

Subjects. Two hundred forty female undergraduate students in introductory psychology at Ohio State University served as Ss. A team consisted of two members, resulting in 120 teams. The Ss were assigned randomly to teams and teams were assigned randomly to experimental groups.

Experimental Task

The experimental paradigm was a standard multiple-cue probabilistic inference task (see Dudycha & Naylor, 1966; Naylor & Schenck, 1968). In this task, Ss are shown a series of two-digit numbers (X_i values or cue values), one at a time, and asked to make predictions (Y_{s_i} values) as to what two-digit criterion number (Y_{e_i} value) is associated with each cue number. After each prediction Ss are shown the "correct" answer. Thus, on each trial

X_i = cue or stimulus value on Trial i

Y_{s_i} = an S's prediction of criterion on Trial i

Y_{e_i} = actual criterion value on Trial i

For any given S the correlational relationship between cue and criterion values is fixed over the n experimental trials—that is, each of the n pairs of numbers

TABLE 1

STIMULUS, CRITERION, AND RESPONSE DATA IN A TWO-CUE, TWO-PERSON TEAM INFERENCE TASK

| Trial | Actual state of nature | | Human performance | | | |
|-------|------------------------|----------------|----------------------------------|-----------------|-------------------------|----------------|
| | Cue variables | | Individual team member responses | | Composite team response | |
| | X ₁ | X ₂ | Y _e | Y _{S1} | Y _{S2} | Y _c |
| 1 | | | | | | |
| 2 | | | | | | |
| 3 | | | | | | |
| 4 | | | | | | |
| 5 | | | | | | |
| 6 | | | | | | |
| ⋮ | | | | | | |
| n | | | | | | |

represents a point from a bivariate normal frequency distribution which describes that specific correlational relationship. Thus *S*'s task is to become what Peterson and Beach (1967) would call an "intuitive statistician" and learn the underlying relationship between *X* and *Y_e*, so that they may predict *Y_e* with increasing accuracy as their experience with the task increases.

Performance Measures

The linear regression model, when applied to a team multiple-cue inference task, provides a number of interesting and powerful indexes of both individual and team performance (for a detailed discussion of these measures see Hursch, Hammond, & Hursch, 1964, and Naylor & Schenck, 1965). Consider a team inference situation involving two cue variables and two team members. Each team member is asked to look at both cues and then make his own prediction concerning the correct answer on that trial. Following this, the two team members are asked to "get together" and come up with a composite or team prediction (*Y_c*). Such a situation is presented in Table 1. Now for each team, let

$$\hat{Y}_e = b_{1e}^* X_1 + b_{2e}^* X_2$$

= optimal linear prediction equation (1)

$$\hat{Y}_c = b_{1c}^* X_1 + b_{2c}^* X_2$$

= policy equation of team (2)

If, after *n* trials, solutions are found for both Equations (1) and (2) and then \hat{Y}_e and \hat{Y}_c values for each of the *n* trials are actually computed in turn, the following team performance measures may formally be defined:³

1. $R_e = r_{Y_e \hat{Y}_e}$ = system ecology
2. $r_a = r_{Y_e \hat{Y}_c}$ = team achievement
3. $R_c = r_{Y_c \hat{Y}_c}$ = team consistency
4. $r_m = r_{\hat{Y}_e \hat{Y}_c}$ = team policy matching

³ All correlations are over *n* experimental trials.

All of the above measures, with the exception of R_e , which defines the system ecology, are indexes describing different aspects of team inference behavior. It is possible to obtain numerous other measures concerning individual team member performance as well, but this paper will examine only team performance under various conditions.

Definition of Variables

The experiment involved the study of both the work structure and task structure variables. However, as was pointed out earlier, task structure is itself a function of three variables—complexity, organization, and redundancy. For purposes of the present study, task structure was formally defined in terms of the multiple regression model as follows⁴:

$$R^2_{Y_e \cdot X_1, X_2, X_1 X_2} = R^2_{Y_e \cdot X_1, X_2} + (R^2_{Y_e X_1 X_2 \cdot X_1, X_2})(1 - R^2_{Y_e \cdot X_1, X_2}). \quad (3)$$

Equation (3) may be expressed in words as Task structure = Task complexity and Task redundancy + Task organization.

Thus in terms of the multiple regression model, task structure is represented by the multiple coefficient of determination which exists when predicting the criterion from both the separate cues and from all combinations of cues. This term may be decomposed algebraically into two subcomponents. The first term on the right side of the equality in (3) reflects the amount of predictability attributed to the cues as individual predictors. It includes both individual cue predictabilities (cue complexity) and the intercorrela-

⁴ This definition of task structure and the resulting method of measuring the relationships existing between or among task components can be generalized by replacing all the multiple and multiple partial correlation coefficients by multiple and multiple partial correlation ratios (Rao, 1965).

TABLE 2
SUMMARY OF EXPERIMENTAL CONDITIONS FOR THE
TWELVE DIFFERENT GROUPS

| Group | Work structure | Task structure | Task organization |
|-------|----------------|----------------|-------------------|
| 1 | Model 1 | .80 | .00 |
| 2 | Model 1 | .40 | .00 |
| 3 | Model 1 | .80 | .40 |
| 4 | Model 1 | .40 | .40 |
| 5 | Model 2 | .80 | .00 |
| 6 | Model 2 | .40 | .00 |
| 7 | Model 2 | .80 | .40 |
| 8 | Model 2 | .40 | .40 |
| 9 | Model 3 | .80 | .00 |
| 10 | Model 3 | .40 | .00 |
| 11 | Model 3 | .80 | .40 |
| 12 | Model 3 | .40 | .40 |

Note.—Each group consists of 10 two-person teams.

tions between cues (cue redundancy). In the two-cue case it may be written

$$R^2_{Y_e \cdot X_1, X_2}$$
$$= \frac{r^2_{Y_e X_1} + r^2_{Y_e X_2} - 2r_{X_1 X_2} r_{Y_e X_1} r_{Y_e X_2}}{1 - r^2_{X_1 X_2}} \quad (4)$$

Equation (4) reflects the general interplay between component redundancy and component complexity. Note that any increase in the redundancy among components detracts from the impact of component complexity. Given that no redundancy exists between task components, Equation 4 reduces to

$$R^2_{Y_e \cdot X_1 X_2} = r^2_{Y_e X_1} + r^2_{Y_e X_2}$$
$$= \text{sum of component complexities} \quad (5)$$

The second term to the right of the equality in Equation (3) represents task organization. It is the squared multiple partial correlation of the criterion with the unique combinations of the cues times the residual of task complexity, that is, the unique contribution to the multiple contributed by the various combinations of cues. As a definition of task organization, the term represents an independent dimension and thus satisfies the demand of the model that organization be a task characteristic which the team members must learn independently of complexity and redundancy. Any value of $R^2_{Y_e X_1 X_2 \cdot X_1, X_2} > 0$ will, of course, result in an increase in task structure.

To summarize, using the linear regression model to define task structure and its determinants results in the following:

- Task structure

Component complexity

Component redundancy
- = total predictability of the criterion using the cues both individually and in combination

= predictability of the individual cues

= cue intercorrelation

Component organization = predictability of the cues when used in combination with each other

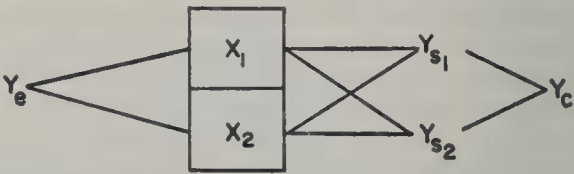
Experimental Design

The design was a $2 \times 2 \times 3 \times 5$ repeated-measures factorial with two levels of task structure (.80 and .40), two levels of task organization (.00 and .40), three levels of work structure, and five blocks of 40 trials each. Task redundancy was kept at .00 (cues were uncorrelated) for all conditions. A summary of the design is given in Table 2.

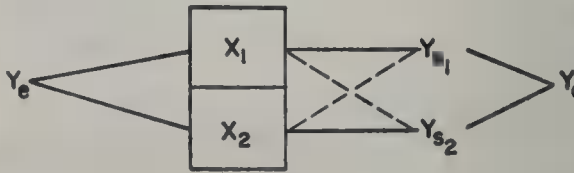
The three work structures differed in the number of cues available as source of information to a team member and/or in how many of these cues Ss were instructed to use. In Model 1, shown in Figure 1, each S had both cues presented and was instructed to predict from both of the sets. In Model 2 both sets were again presented, but Ss were instructed to predict from only one of the sets. Thus, one team member predicted using the X_1 cue and the other using the X_2 cue, although both cues were visible to each S. In Model 3 each team member had only one of the cues available for prediction.

To create task stimuli which would satisfy the specified values for task structure (.80, .40) and task organization (.00, .40) and would keep task redundancy constant at .00, three sets of 200 uncorrelated z scores were generated through the use of a computer program

Model 1



Model 2



Model 3

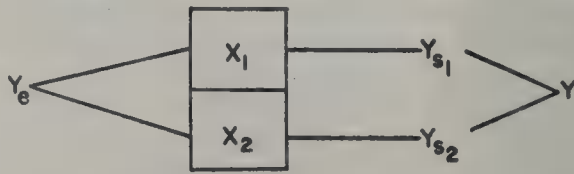


FIG. 1. Two-member team work structures.

(Wherry, Naylor, Wherry, & Fallis, 1965). Two of the sets were multiplied together to obtain an interaction or cross-product term. Each of the numbers in the four sets was then multiplied by the desired regression coefficient and summed to obtain the criterion values (see Table 3). Using the above procedure, 200 criterion numbers were generated from each equation whose multiple correlation with all four sets, theoretically, was unity, and whose correlation with each cue set was equal to the respective regression coefficient. Due to rounding errors, however, the desired empirical values deviated from those obtained, as shown in Table 3.

Two of the sets, X_1 and X_2 , were the actual cues presented to Ss. In all the defining equations X_1 and X_2 were weighted equally and the sum of their predictabilities of the criterion represented the task complexity. The interaction term's (X_1X_2) predictability of the criterion defined the level of task organization, and the fourth set, X_3 , served as an error term. The sum of X_1 , X_2 , and X_1X_2 predictabilities represented the total amount of task predictability, that is, task structure.

Experimental Procedure

A group testing procedure was employed. All 10 teams for a particular experimental condition were tested in the same 2-hr. session. A session involved 200 trials, where a trial consisted of (a) displaying the cue(s), (b) getting individual S predictions of the criterion, (c) getting a team prediction, and (d) displaying the correct answer for that trial.

The Ss were instructed that each was to make an individual prediction of the value of the criterion from

the cue(s) available to them and to record their predictions on an answer sheet. Following their individual predictions, members were then to "get together" with their partners and make a composite or team prediction on the basis of each partner's individual prediction. The Ss were told that their task was to predict the criterion as accurately as possible using the cue(s), and that the cue(s) and criterion were related. They were told not to expect to predict perfectly, but that with practice they should be able to increase in skill. Cue sets were presented to Ss on stapled sheets, while the criterion numbers were visually displayed from an opaque projector. The task was group paced—the teams were allowed approximately 40 sec. to make predictions and then the criterion was displayed for 5 sec. before E proceeded to the next trial. Three sample trials were given prior to the start of the actual testing session, and a 5-min. rest period was introduced in all sessions at the end of 100 trials.

RESULTS

A multiple regression analysis of the form

$$\hat{Y}_e = a + b_{1c}X_1 + b_{2c}X_2 + b_{12c}X_1X_2 \quad (6)$$

was computed for the 200 team responses for each of the 120 teams. This equation was then used to obtain \hat{Y}_e values on each of the 200 trials for that team. Since Equation (6) represents a team's prediction "strategy," the \hat{Y}_e values represent what the team would have predicted had they used that strategy perfectly (without error) during the testing session.

A similar analysis was run on the criterion data for each group. This equation was of the form

$$\hat{Y}_e = a + b_{1e}X_1 + b_{2e}X_2 + b_{12e}X_1X_2 \quad (7)$$

and represented the optimal prediction strategy for that experimental condition. The \hat{Y}_e values were then computed for each of the 200 trials in that experimental condition—these represented the optimal prediction for that trial. Once these values were computed there were four scores available for each S on each of the 200 trials— Y_e , Y_s , \hat{Y}_e , \hat{Y}_s . These were used to obtain the performance measures described in the previous section. This was done separately for each block of 40 trials for each team. All correlations were transformed to Fisher z values for use in the analyses of variance.

Team Achievement

Significant main effects on team achievement were found due to Task structure

TABLE 3

THEORETICAL VALUES USED IN GENERATING CUES AND THE RESULTING EMPIRICAL RELATIONSHIPS

| Criteria generation equations | | | | | |
|--|--------------|--------------|-----------------|------------------------------------|-------------|
| $Y_{e_1} = .632$ | $X_1 + .632$ | $X_2 + .000$ | $X_1X_2 + .447$ | X_3 | |
| $Y_{e_2} = .447$ | $X_1 + .447$ | $X_2 + .447$ | $X_1X_2 + .632$ | X_3 | |
| $Y_{e_3} = .447$ | $X_1 + .447$ | $X_2 + .000$ | $X_1X_2 + .775$ | X_3 | |
| $Y_{e_4} = .000$ | $X_1 + .000$ | $X_2 + .632$ | $X_1X_2 + .775$ | X_3 | |
| Empirical correlations of cues with criteria | | | | | |
| Criterion | X_1 | X_2 | X_1X_2 | Task predictability (r_e^2) | |
| | | | | Empirical | Theoretical |
| Y_{e_1} | .617 | .630 | -.083 | .781 | .800 |
| Y_{e_2} | .378 | .482 | .596 | .754 | .800 |
| Y_{e_3} | .442 | .417 | -.111 | .377 | .400 |
| Y_{e_4} | -.041 | .009 | .608 | .370 | .400 |

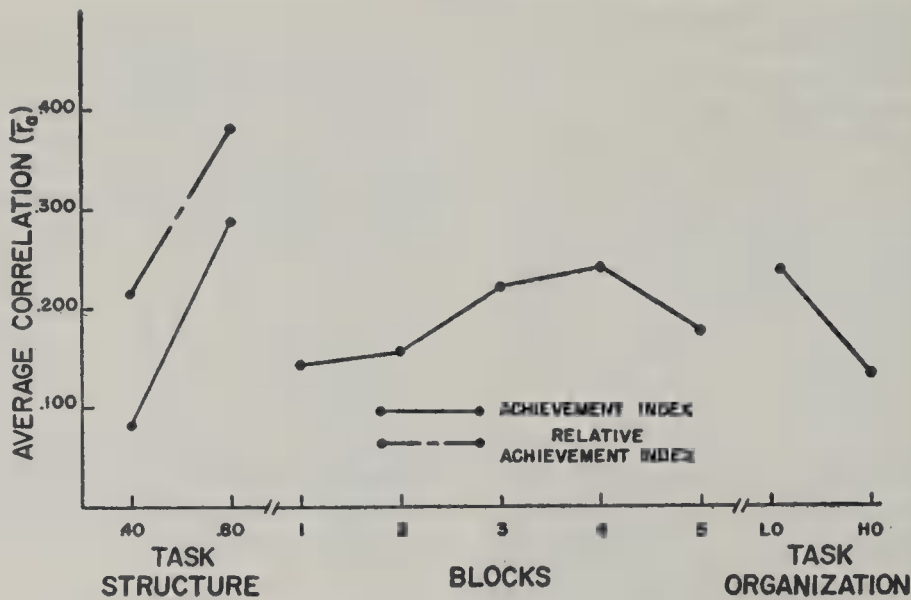


FIG. 2. Main effects from analysis of variance performed on team achievement (r_a).

($F=45.13$; $df=1/108$; $p<.01$), Task organization ($F=11.86$; $df=1/108$; $p<.01$), and Blocks ($F=6.34$; $df=4/432$; $p<.01$). There was also evidence for a Blocks \times Task structure interaction ($F=3.21$; $df=4/432$; $p<.01$).

The three main effects are shown in Figure 2. Greater absolute achievement occurred under conditions of greater task structure; that is, those teams performing in an ecology which was 80% deterministic did better than those teams operating in an ecology which was only 40% deterministic. Absolute achievement, however, may not be the best comparison to employ between these two environments. Instead of examining absolute achievement it may be more appropriate to examine achievement relative to maximum possible achievement. It can be demonstrated algebraically that R_e represents an upper bound limit on achievement in a linear prediction task. Thus, the ratio r_a/R_e may be taken as a measure of relative achievement.

With this measure it can be asked whether teams having more structured tasks do better relative to theoretical maximum performance than do teams having less structured tasks. Figure 2 also shows these data—the answer is obviously the same as it was with the absolute achievement measure.

The significant task organization effect was due to a sharp decrement in team achievement

under conditions of high task organization. The presence of interrelationships among the task components thus contributed little to achievement, and proved to be a somewhat complex task property.

The significant Blocks \times Task structure interaction was examined and found to be a result of more rapid learning for teams operating in tasks of greater structure. Indeed, when the interaction means were analyzed using a Newman-Keuls procedure for ordered means, no learning trend was found for teams having low task structure. Apparently these teams reached a performance asymptote by the end of the first block of 40 trials and did not improve thereafter, while teams under high task structure continued to improve through the first four blocks. A small but significant decrement was observed for performance in Block 5. This decrement has been observed before in tasks of this type (Dudycha & Naylor, 1966; Naylor & Schenck, 1968) and appears to be a motivational effect due to boredom or fatigue.

Team Consistency

Only one significant main effect was observed in the analysis of team consistency data. This was due to Task Structure ($F=16.90$; $df=1/108$; $p<.01$). This effect is shown in Figure 3.

Note that the teams having the most structured tasks responded to the situation by

performing in a more consistent fashion. These teams adopted prediction strategies which were substantially more stable than those adopted by teams operating in task structures which were much less deterministic. This finding is also compatible with prior research on individual prediction behavior (e.g., see Dudycha & Naylor, 1966; Uhl, 1963) which has demonstrated that individual performance consistency increases as the structure of the environment increases. Also worthy of mention is the fact that relative to the level of task structure (R_e), teams under the low structure condition actually were more consistent than were the teams under the high structure condition (the ratio r_s/R_e is shown in Figure 3). This finding is also congruent with results obtained on individual prediction behavior. Thus, a decrease in task structure does not, apparently, result in a proportional decrease in the consistency of the prediction strategy employed by the teams.

Team Strategy Matching

The matching index, r_m , reflects the degree to which the team's best fitting policy equation or strategy "matches" the optimal prediction strategy. Thus, in a sense, it reflects the degree to which the team is performing in a truly optimal manner. Significant effects on matching behavior were observed due to Task

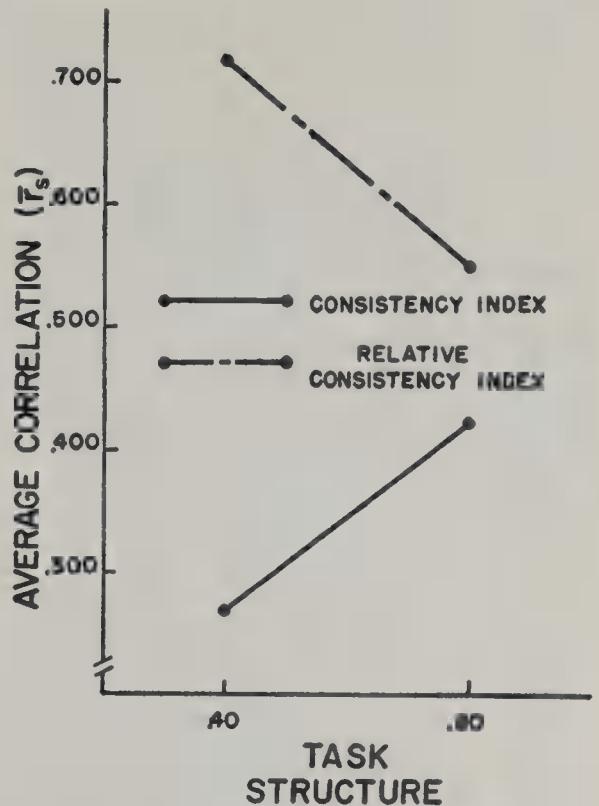


FIG. 3. Task structure main effect from analysis of variance performed on team consistency (r_s).

structure ($F = 15.42$; $df = 1/108$; $p < .01$), Task organization ($F = 33.92$; $df = 1/108$; $p < .01$), and Blocks ($F = 35.10$; $df = 4/432$; $p < .01$). Several interactions with blocks were also significant—Blocks \times Task struc-

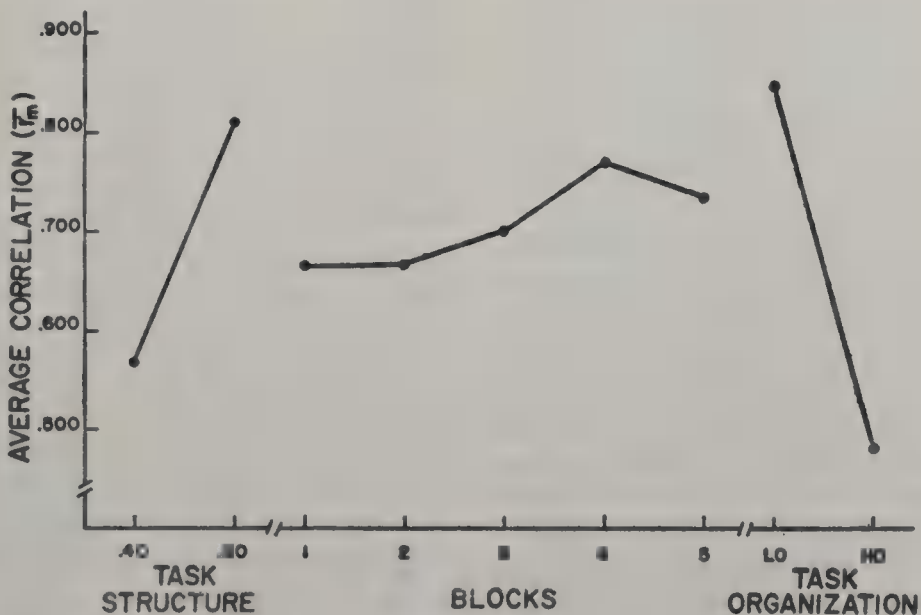


FIG. 4. Main effects from analysis of variance performed on team matching (r_m).

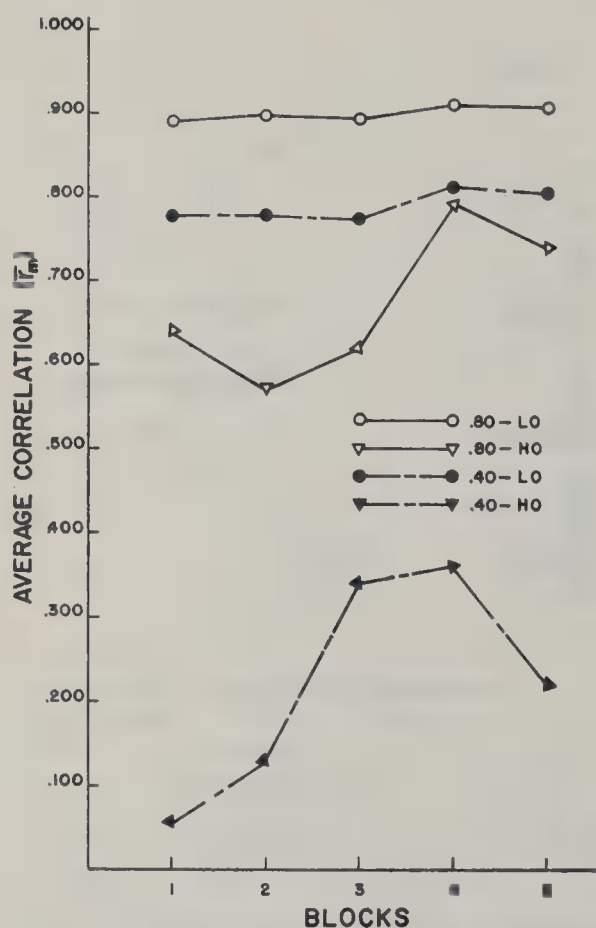


FIG. 5. Blocks \times Organization \times Task structure interaction from analysis of variance performed on team matching (r_m).

ture ($F = 6.07$; $df = 4/432$; $p < .01$), Blocks \times Organization ($F = 8.58$; $df = 4/432$; $p < .01$), and Blocks \times Organization \times Task structure ($F = 5.50$; $df = 4/432$; $p < .01$). The significant main effects are shown in Figure 4.

More efficient matching was obtained under high task structure conditions and also with tasks having low organization. There was a substantial learning effect over blocks with the exception of Block 5 which showed a decrement in performance similar to that observed with achievement. Apparently, then, a highly structured environment permits a team not only to be more consistent in its strategy but also to develop a more appropriate response strategy—appropriate in the sense that it more closely approximates an optimal strategy. Task organization, on the other hand, detracts from the ability of teams to develop appropriate strategies. Since task organization is

defined in terms of interrelationships between or among task components, it would appear that when such interrelationships exist the difficulty of developing an optimal response strategy is substantially increased.

The significant interactions with blocks were all indicative of differential learning rates under various conditions. Some were of considerable interest. For example, the Blocks \times Organization interaction was a result of a noticeable learning trend during the last several blocks for the high organization groups as compared to virtually no learning being observed for the low organization groups. These latter groups had r_m values near .90 even during the first block and showed little increase over that value throughout the remaining trials. The most logical interpretation of this interaction is that (a) strategies based solely on the individual cues are adopted very quickly by teams and once adopted are systematically maintained with little change and (b) strategy sophistication based upon complex interrelationships between cues occurs much later in the task.

This interpretation was further supported by the Blocks \times Organization \times Task structure interaction (see Figure 5). Groups having no task organization component showed little increase in matching ability across blocks, regardless of the level of task structure (r_m was consistently quite high for these groups). For groups having a high task organization component considerable improvement in matching occurred, with the greatest rate of improvement being observed with the groups working under low task structure. In this latter condition, task organization accounted for all of the total task structure and indeed, if any matching was to occur, it had to be on the basis of the organizational component of the task. Thus, the task organization concept is clearly a difficult one for teams to acquire—much more difficult than the more straightforward individual cue-criterion relationships. However, it can be learned and the degree to which it is learned appears to be related to the degree to which it contributes to total task structure. Also, task organization appears to be learned only after the more simple relationships are mastered.

Predicting Team Responses from Member Responses

The team consistency measure discussed earlier represents the degree to which a team's behavior may be predicted from the task components (cues) and their interrelationships. It seemed desirable also to examine the question of the degree to which team behavior could be predicted from a knowledge of individual member behavior. To accomplish this, individual member predictions were used as predictors of the team response in a multiple regression analysis. Thus, there were two predictors (two team members) which were regressed on the criterion of the composite team response. This was done for every team in every condition. The resulting multiple R 's were transformed into Fisher z values and analyzed using an ANOVA design identical to that used with the previously mentioned performance measures.

Two significant effects were found. Work structure was significant ($F=4.80$; $df=2/108$; $p<.05$) as was the Blocks \times Task organization \times Task structure interaction. Team responses in Model 1 were most predictable from member responses ($\bar{R} = .669$), team responses in Model 2 were least predictable ($\bar{R} = .528$), and Model 3 led to intermediate predictability ($\bar{R} = .606$). A Newman-Keuls test revealed the Model 1 versus Model 2 comparison to be the only significant difference among the three means. An examination of the interaction effect indicated no easily interpretable patterns.

DISCUSSION

While the findings demonstrating the effect of both task structure and task organization upon team performance were not surprising, it was unexpected to see the lack of influence exerted by the work structure variable. On an intuitive basis, one would anticipate the different work structures might, at the very least, comprise a very critical variable in those cases where task organization was high and perhaps a much less important variable under low task organization—however, no such interaction appeared. Thus, even in those cases where each S was able to see both cues, there was no facilitating effect regarding sensitivity to the

interrelationship which existed between cues in the high organization experimental conditions.

The fact that work structure failed to influence any of the three team performance measures—achievement, consistency, and matching—might, at first glance, appear to be in direct contradiction to the research of Leavitt (1951), Guetzkow and Simon (1955), Shaw (1954a, 1954b), Shaw and Blum (1965), and Faucheux and MacKenzie (1966). However, this difference is more apparent than real. In all of these studies the concept of group (or team) structure was defined in terms of communication network arrangements between individual team members, not in terms of the way in which individuals were formally assigned to subtask roles. Thus, these earlier studies were either manipulating or examining post hoc the influence of what Dickinson and Naylor (1966) would define as communication structure rather than work structure. The logical separation of these two types of structure strikes the authors as being of critical importance. For example, one logical inference arising out of the comparison of the results of the present study to those of prior researchers is that communication structure may indeed be a much more critical variable in determining team performance than is work structure. Support for this notion can be found in the study of Williges, Johnston, and Briggs (1966) who found no significant effect on performance due to training conditions (a manipulation that would be classed under work structure in the Dickinson-Naylor paradigm) but which did find a significant effect due to communication category. However, neither that study nor the present research provides a clear test of the relative importance of these two types of structure. The implication, nevertheless, seems quite clear and quite important for team performance.

To accuse the work structure parameter as being completely without effect is not just, since the analysis using individual team member responses as predictors of composite team responses did show an effect due to the three different models. The composite team decision was most related to the predictions of the individual team members in Model 1. Since both team members had access to and were

asked to use the same information in this model, team member agreement should be highest in Model 1 and therefore (a) there should be less conflict in making a composite decision and (b) that decision should be fairly compatible with each team member's own view. What was particularly interesting, however, was the finding that Model 2 team responses were less predictable from its member predictions than were the composite responses in Model 3. In Model 2 each team member had access to the information being used by his partner but was instructed not to use this information in making his own predictions. Thus, as in Model 1 a team member is able to learn something about his partner's response system, that is, the process of interpersonal learning can take place in Model 2 just as easily as in Model 1. In Model 3, however, there is absolutely no opportunity for a team member to learn anything about his partner's response system nor can he use the information in the second cue in his own predictions since he has access to only one cue. Yet one finds that team responses are more systematically related to individual member predictions in Model 3 than in Model 2. This implies that team members in Model 2 had greater difficulty in arriving at a joint decision compatible with their individual decision than did teams in Model 3. This result appears directly in opposition to the notion in Hammond, Wilkins, and Todd (1966) that interpersonal conflict should be reduced in situations where knowledge of how the other person is interacting with the environment is available to S.

The most potent variable influencing team performance was that of task structure. All three performance measures were related to this variable. The more structured the task, the higher was team achievement, team consistency, and team matching. These data were all consistent with previous findings on individual performance in multiple-cue inference situations (e.g., see Naylor & Schenck, 1968). Also consistent with prior research was the finding that team achievement relative to maximum possible achievement tended to increase as task structure increased and that team consistency relative to environmental consistency tended to decrease as task struc-

ture increased. This latter result—a consistent finding in individual inference research—is indicative of the unwillingness of teams to perform in a random fashion (have a random team strategy) even when they are presented with a task environment that is close to being a random environment (the low task structure conditions). Apparently teams, like individuals, will tend to adopt a systematic strategy even when that strategy can be of no earthly use to them as far as the task is concerned.

The task organization variable, even though its presence added to the overall predictability of the task, clearly resulted in a decrement in the ability of teams to achieve and to develop an optimal prediction strategy (match). This seems compatible with the research of Briggs and Waters (1958) who found that, for individual Ss, tasks having interrelationships among task components were more difficult than tasks in which the components were independent. Also, the inference studies dealing with the ability of Ss to learn higher order stimulus-criterion relationships indicate the high difficulty level of this kind of task characteristic (Hammond & Summers, 1965; Summers & Hammond, 1966). As mentioned earlier, however, it was unexpected to find that, in the team inference situation, work structure failed to become a moderator of this outcome.

Finally, the finding that the learning of strategies by teams appears to "progress" from individual cue-criterion relationships (task complexity) to the more complex interaction relationships (task organization) supports Fuch's (1962) "progression hypothesis" concerning the acquisition of skilled performance. Teams only began to acquire some skill with the organization aspect of the task after they had substantially mastered task complexity (individual cue-criterion relationships). This agrees with Fuch's hypothesis which holds that more basic (zero order) functions are learned first and that learning then progresses in a hierarchical fashion up through more complex (higher order) aspects of the task.

REFERENCES

- BRIGGS, G. E., & WATERS, L. K. Training and transfer as a function of component interaction. *Journal of Experimental Psychology*, 1958, 56, 492-500.

- DICKINSON, T. L., & NAYLOR, J. C. Effects of task complexity and organization on team behavior using the multiple regression model. Paper presented at Midwestern Psychological Association, May 1966.
- DUDYCHA, L., & NAYLOR, J. C. Characteristics of the human inference process in complex choice behavior situations. *Organizational Behavior and Human Performance*, 1966, 1, 110-128.
- FAUCHEUX, C., & MACKENZIE, K. D. Task dependency of organizational centrality and its behavioral consequences. *Journal of Experimental Social Psychology*, 1966, 2, 361-375.
- FUCHS, A. H. The progression-regression hypothesis in perceptual-motor skill learning. *Journal of Experimental Psychology*, 1962, 63, 177-182.
- GUETZKOW, H., & SIMON, H. A. The impact of certain communication nets upon organization and performance in task-oriented groups. *Management Science*, 1955, 1, 233-250.
- HAMMOND, K. R., & SUMMERS, D. A. Cognitive dependence on linear and nonlinear cues. *Psychological Review*, 1965, 72, 215-224.
- HAMMOND, K. R., WILKINS, M. M., & TODD, F. J. A research paradigm for the study of interpersonal learning. *Psychological Bulletin*, 1966, 4, 221-232.
- HURSCHE, C. J., HAMMOND, K. R., & HURSCHE, J. L. Some methodological considerations in multiple-cue probability studies. *Psychological Review*, 1964, 71, 42-60.
- LEAVITT, H. J. Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology*, 1951, 46, 38-50.
- NAYLOR, J. C., & BRIGGS, G. E. Team-training effectiveness under various conditions. *Journal of Applied Psychology*, 1965, 49, 223-229.
- NAYLOR, J. C., & SCHENCK, E. A. A revised summary of the multiple cue model and its related performance indices. Mimeo, Ohio State University, 1965.
- NAYLOR, J. C., & SCHENCK, E. A. The influence of cue redundancy upon the human inference process for tasks of varying degrees of predictability. *Organizational Behavior and Human Performance*, 1968, 3, 47-61.
- PETERSON, C. R., & BEACH, L. R. Man as an intuitive statistician. *Psychological Bulletin*, 1967, 68, 29-46.
- RAO, C. R. *Linear statistical inference and its applications*. New York: Wiley, 1965.
- SHAW, M. E. Some effects of problem complexity upon problem solution efficiency in different communication nets. *Journal of Experimental Psychology*, 1954, 48, 211-217. (a)
- SHAW, M. E. Some effects of unequal distribution of information upon group performance in various communication nets. *Journal of Abnormal and Social Psychology*, 1954, 49, 547-553. (b)
- SHAW, M. E., & BLUM, J. M. Group performance as a function of task difficulty and the group's awareness of member satisfaction. *Journal of Applied Psychology*, 1965, 49, 151-154.
- SUMMERS, D. A., & HAMMOND, K. R. Inference behavior in multiple-cue tasks involving both linear and nonlinear relations. *Journal of Experimental Psychology*, 1966, 71, 751-757.
- UHL, C. N. Learning of interval concepts. I. Effects of differences in stimulus weights. *Journal of Experimental Psychology*, 1963, 66, 264-273.
- WHERRY, R. J., SR., NAYLOR, J. C., WHERRY, R. J., JR., & FALLIS, R. F. Generating multiple samples of multivariate data with arbitrary population parameters. *Psychometrika*, 1965, 30, 303-313.
- WILLIGES, R. C., JOHNSTON, W. A., & BRIGGS, G. E. Role of verbal communication in teamwork. *Journal of Applied Psychology*, 1966, 50, 473-478.

(Received May 10, 1968)

MOTIVATION OF RESEARCH AND DEVELOPMENT ENTREPRENEURS:

DETERMINANTS OF COMPANY SUCCESS¹

HERBERT A. WAINER AND IRWIN M. RUBIN²

Massachusetts Institute of Technology, Sloan School of Management

Fifty-one technical entrepreneurs were studied, focusing upon the relationships between their motivation and company performance. More specifically, the relationships between the entrepreneurs' need for achievement, need for power, and need for affiliation were related to the performance of the 51 small companies they founded and operated. The results indicate that high need for achievement and moderate need for power are associated with high company performance. The effects of need for power and need for affiliation on performance seem to be derived through their influence on leadership styles.

In an attempt to associate need for achievement (*n Ach*) and economic development, McClelland (1961) looks to the entrepreneur as the one who translates *n Ach* into economic development. The entrepreneur in McClelland's scheme is "the man who organizes the firm (the business unit) and/or increases its productive capacity [p. 205]."

The present authors' aim was to test McClelland's macro theory of economic growth at the micro level of organizational performance. The principle interest in considering McClelland's work stems from his discussions of who entrepreneurs are and of their different behavioral styles predicted from differences in need patterns. McClelland's underlying assumption is that entrepreneurs have a high *n Ach* and that in business situations this

high *n Ach* will lead them to behave in certain ways and have certain tendencies.

Based on McClelland's discussion, the present authors raised the proposition that the degree to which an entrepreneur is motivated by *n Ach* directly influences his skill as an entrepreneur and consequently his enterprise's performance. The major hypothesis to be tested concerns the relationship between an entrepreneur's level of *n Ach* and his company's performance.

Schrage (1965), in testing the relationship between the entrepreneur's *n Ach* and company performance, reported that companies run by entrepreneurs who have a high *n Ach* tend to have either high profits or losses ($\pm 3\%$ of sales), while those run by low *n Ach* entrepreneurs tend to have low profits or losses ($\leq 3\%$ of sales). Reanalysis of his data by the present authors sheds considerable doubt on the validity of his findings. The primary source of doubt was a discrepancy between the scores Schrage used for *n Ach* and those subsequently derived when the same protocols were rescored by the Motivation Research Group at Harvard. The fact that his results departed markedly from established theory further substantiates this concern.

In addition to the relationship between *n Ach* and company performance, the authors were interested in the interrelationships among three needs, *n Ach*, need for power (*n Pow*), and the need for affiliation (*n Aff*), with respect to company performance. *n Pow* is defined by Atkinson (1958) as "that dis-

¹ The research presented in this paper was supported in part by grants from the Massachusetts Institute of Technology Center for Space Research and by the National Aeronautics and Space Administration (NsG-235 and NsG-496). However the findings and views reported are those of the authors and do not necessarily reflect those of the supporting agencies. This work was done in part at the Massachusetts Institute of Technology Computation Center.

The authors wish to acknowledge the work done by Charles W. McLaughlin, a master's degree candidate in the Massachusetts Institute of Technology, Sloan School of Management, in the collection of data for this paper.

² Requests for reprints should be sent to Irwin M. Rubin, Room 52-538, Massachusetts Institute of Technology, Sloan School of Management, 50 Memorial Drive, Cambridge, Massachusetts 02139.

position, directing behavior toward satisfactions contingent upon the control of the means of influencing another person [p. 105]."

n Aff is concerned with the establishment, maintenance, or restoration of positive affective relationships with other people, that is, friendships. Statements of liking or desire to be liked, accepted, or forgiven are manifestations of this motive (Atkinson, 1958). McClelland's (1961) discussion of the joint product of *n* Pow and *n* Aff in relation to dictatorship stimulated this aspect of the inquiry. He found that *n* Pow was not related to economic growth but was related to style of leadership. More specifically, the combination of a high *n* Pow and a low *n* Aff was associated with the tendency of a country to resort to totalitarian methods as a style of leadership.

The present authors propose that *n* Ach has behavioral manifestations different than either *n* Pow or *n* Aff in terms of the individual's relationships with people. *n* Pow and *n* Aff are interpersonally oriented needs. Implicit in their definitions is the existence of other human beings whom the *n* Pow or *n* Aff motivated individual can influence and control, or with whom he can be friends. *n* Ach, on the other hand, seems to be a more internalized need. The *n* Ach motivated individual may need other people to help him satisfy his *n* Ach, but the nature of his relationship with them, or more appropriately his effectiveness with them, will be determined by other needs. The authors suggest that *n* Ach is a primary consideration determining noninterpersonally related behavior that leads to high company performance. *n* Pow and *n* Aff are primary considerations determining interpersonal behavior that affects company performance. *n* Pow and *n* Aff, then, can be looked upon as having strong implications as determinants of management style.

Numerous other attempts have been made to identify those personality traits which differentiate leaders from nonleaders or effective leaders from ineffective leaders. These studies have, in general, failed to find any consistent pattern of differentiating traits. In a broad sense, the present research is analogous to these prior efforts in that it seeks

to explain company performance on the basis of certain personality characteristics of the president. Steps were taken, however, in anticipation of two potential problem areas: (a) that personality description and measurement themselves are not yet adequate; (b) that the groups studied have usually been markedly different from one another and this may have concealed a relation between personality and the exercise of leadership that would have appeared within a more homogeneous set of groups or situations.

The major personality variable of interest in the present study is the need for achievement. On the basis of the existing body of research, McClelland's version of the Thematic Apperception Test (TAT) was deemed a reliable means of measuring *n* Ach (Atkinson, 1958; McClelland, 1961). With respect to the second problem area, a very homogeneous set of groups has been examined, thus mitigating the potential influence of the "situation."

For these reasons, the focus in this study was upon the new, small, technically based enterprise. The entrepreneur president of such a company has placed himself in a situation where his *n* Ach, to the extent that it exists, can readily be translated into concrete behavior. He starts the company, hires the people, and motivates them, sells, plans, takes risks, and so on. It is his personality and motivation that mold the company in its every aspect. Furthermore, in such situations, the entrepreneur's efforts and decisions are likely to be very important in determining the initial success of the venture.

METHOD

Fifty-one small technically based companies in the Boston area comprised the sample. All were at least 4 but less than 10 yr. old at the time of the study and all were "spin-offs" from one of the Massachusetts Institute of Technology research laboratories or industrial laboratories around the Boston area. They ranged in business activities from service, such as computer software development, to manufacturing, such as special purpose computers and welded modules. Company and entrepreneurial personality information were gathered from the entrepreneur president. The typical entrepreneur, based on the central tendencies for the total sample of entrepreneurs, was approximately 36 yr. of age when he started his new enterprise, was educated to the master's degree level, and had considerable

TABLE 1

MEANS, MEDIAN, AND RANGES OF
VARIABLES MEASURED

| Variable | <i>M</i> | <i>Mdn</i> | Range |
|-------------|----------|------------|-------------|
| n Ach | 5.9 | 5.0 | -5 to 18 |
| n Pow | 9.7 | 9.5 | 0 to 19 |
| n Aff | 3.5 | 3.0 | 0 to 16 |
| Growth rate | .40 | .375 | 0.0 to 2.10 |

experience at a technically advanced research laboratory prior to starting his new enterprise. Among the information gathered were company yearly sales figures and scores on McClelland's version of the TAT for each entrepreneur. The yearly sales figures were used as the basis for determining the growth rate, defined in detail below. The index of performance was derived from the growth rate. The TATs were scored for n Ach, n Pow, and n Aff by the Motivation Research Group at Harvard University. The resulting scores were the basis for analysis of the strength of various needs in relation to performance.³

Growth rate is defined as follows: *annual increase in the logarithm of sales volume between the second and most recent year reported*. For example, Company A is 7 yr. old. Its second-year sales were \$100,000 and its last year (seventh) sales were \$950,000. These two sales values are plotted on semi-log paper. The growth rate is indicated by the percent rate of change from year to year. This is, of course, constant over the 7 yr. The growth rate in this case would be approximately .56. Table 1 summarizes the general characteristics of the four variables with which this paper is concerned.

The method of analysis in all cases was a comparison of high, moderate, and low groups. Equality of sample size, within the limits of tied observations, was the criterion used in making these groupings. Standard correlational techniques were feasible in many cases, and, where appropriate, coefficients are presented in footnotes. However, since such techniques often mask nonlinear trends in relationships, the Mann-Whitney *U* test, one of the most powerful of the nonparametric statistical tests, was used. Furthermore, correlation techniques focused on differences between two variables based on individual differences from case to case. On the other hand, the Mann-Whitney *U*, a difference in medians test, analyzes differences between characteristics of groups of data. The authors feel that TAT scoring procedures are not yet precise enough to enable researchers to use individual differences as the basis for comparison.

³ Average intercoder reliabilities of scores from the Motivation Research Group are in the high .80 range.

RESULTS

Analyses of the relationship between the three needs, n Ach, n Pow, and n Aff, and their relation to company performance are presented in this section. In addition, some exploratory results will be presented that focus on the question: Is there a pattern or combination of needs which are related to high company performance? In other words, one set of analyses will focus on the direct relationship between performance and varying degrees of strength in a single need, while a secondary focus will explore effects of several needs taken together on company performance.

Relationship between the Three Needs

The data in Table 2 suggest that, within this sample, n Ach, n Pow, and n Aff are not completely independent.⁴ n Ach appears to be positively related to n Pow and negatively related to n Aff, while n Pow is negatively related to n Aff. It is important to note, however, that in all cases the relationship is nonlinear. In the case of n Ach versus n Pow, for example, only the low n Ach group has a significantly different n Pow score. No differences in n Pow are observed when a comparison of the high versus moderate n Ach groups is made. A similar phenomenon is present in each relationship. In other words, the correlation coefficients reported in Footnote 4 are heavily influenced by a small subset of the total distribution of need scores.

With these qualifications in mind, it is concluded that the three needs are moderately related. Where the relationship between each need and company performance is examined, an attempt will be made to take into account this lack of independence.

Need Strength versus Company Performance

The major hypothesis in this study predicts a direct and positive relationship between an entrepreneur's n Ach and the performance of his company. No directional hy-

⁴ The following are the Kendall Tau correlations between three needs (two-tailed test). n Ach versus n Pow: *T* = .370, *p* < .01, *N* = 51. n Ach versus n Aff: *T* = -.259, *p* < .01, *N* = 51. n Aff versus n Pow: *T* = -.233, *p* < .05, *N* = 51.

potheses were specified concerning the relationships between *n* Pow, *n* Aff, and company performance.⁵

Referring to Table 3, it can be seen that, within the range of moderate to high *n* Ach, a very marked positive relationship exists between *n* Ach and company performance. The growth rate of those companies led by entrepreneurs with a high *n* Ach was almost 250% higher (.73 versus .21) than those companies led by entrepreneurs with a moderate *n* Ach. Here again, however, the relationship is not purely linear since the low *n* Ach group has a mean performance score slightly *higher* than the moderate *n* Ach group but still significantly lower than high *n* Ach group.

n Pow, as can be seen from Table 3, is completely unrelated to company performance. *n* Aff, on the other hand, exhibits a mildly negative, nonlinear, relationship to company performance. The data were then examined to see if the observed relationship between *n* Ach and *n* Aff influenced the relationship found between *n* Ach and performance. No such contamination was found. Of those who were classified in the low *n* Aff group (*n* = 13), only six fell into the high

⁵ The following are the Kendall Tau correlations between the three needs and company performance (growth rate). *n* Ach versus performance: *T* = .15, *p* < .08, *N* = 51 (one-tailed). *n* Pow versus performance: *T* = .05, *p* < .64, *N* = 51. *n* Aff versus performance: *T* = -.11, *p* < .28, *N* = 51.

TABLE 2
RELATIONSHIP BETWEEN THE THREE NEEDS

| Need | Group | | | | | |
|--------------|----------|----------|-------------------|----------|---------|----------|
| | High | | Moderate | | Low | |
| <i>n</i> Ach | | | | | | |
| | ≥ 9 | <i>n</i> | 4 ≥ <i>X</i> ≤ 8 | <i>n</i> | ≤ 3 | <i>n</i> |
| <i>n</i> Pow | 11.3 (A) | 14 | 11.1 (B) | 19 | 6.8 (C) | 18 |
| <i>n</i> Aff | 2.0 (D) | 14 | 3.9 (E) | 19 | 4.4 (F) | 18 |
| <i>n</i> Pow | | | | | | |
| | ≥ 13 | <i>n</i> | 8 ≥ <i>X</i> ≤ 12 | <i>n</i> | ≤ 7 | <i>n</i> |
| <i>n</i> Aff | 2.8 (G) | 15 | 3.0 (H) | 19 | 4.7 (I) | 17 |

Note.—Mann-Whitney *U* test results: A versus B, *p* < .60; A versus C, *p* < .003; B versus C, *p* < .007; D versus E, *p* < .13; D versus F, *p* < .02; E versus F, *p* < .45; G versus H, *p* < .60; G versus I, *p* < .11; and H versus I, *p* < .09.

TABLE 3

RELATIONSHIP BETWEEN *n* Ach, *n* Pow, AND
n Aff AND GROWTH RATE

| Need | Code for Mann-Whitney <i>U</i> results ^a | Strength | Mean growth rate |
|--------------|---|---|------------------|
| <i>n</i> Ach | A | High (≥ 9) <i>N</i> = 14 | .73 |
| | B | Moderate (4 ≥ <i>X</i> ≤ 8) <i>N</i> = 19 | .21 |
| | C | Low (≤ 3) <i>N</i> = 18 | .36 |
| <i>n</i> Pow | A | High (≥ 13) <i>N</i> = 15 | .38 |
| | B | Moderate (8 ≥ <i>X</i> ≤ 12) <i>N</i> = 19 | .47 |
| | C | Low (≤ 7) <i>N</i> = 17 | .36 |
| <i>n</i> Aff | A | High (≥ 4) <i>N</i> = 20 | .33 |
| | B | Moderate (2 ≥ <i>X</i> ≤ 3) <i>N</i> = 18 | .30 |
| | C | Low (≤ 1) <i>N</i> = 13 | .67 |

^a Results of Mann-Whitney *U* tests: *n* Ach versus growth rate: A versus B, *p* < .0001; A versus C, *p* < .006; B versus C, *p* < .08, one-tailed. *n* Pow versus Growth rate: A versus B, *p* < .80; A versus C, *p* < .90; B versus C, *p* < .80, two-tailed. *n* Aff versus growth rate: A versus B, *p* < .81; A versus C, *p* < .16; B versus C, *p* < .10, two-tailed.

n Ach group. *n* Ach, in other words, directly affects company performance, independent of its relationship to *n* Aff.

The results of this section are summarized graphically in Figure 1. The percentage of companies within each subgroup (high, moderate, low), whose performance is above that of the median for the total sample of entrepreneurs, is plotted for each of the needs. Seventy-nine percent of those companies led by entrepreneurs whose *n* Ach was high had a growth rate which was above the median for the total sample of entrepreneurs.

Joint Products of Needs versus Performance

The previous section focused on variations in company performance resulting from each of the three needs (*n* Ach, *n* Pow, and *n* Aff) taken singularly. The aim in this section is to explore the question of whether or not any *pattern* of need strengths appears to be associated with high company performance. In

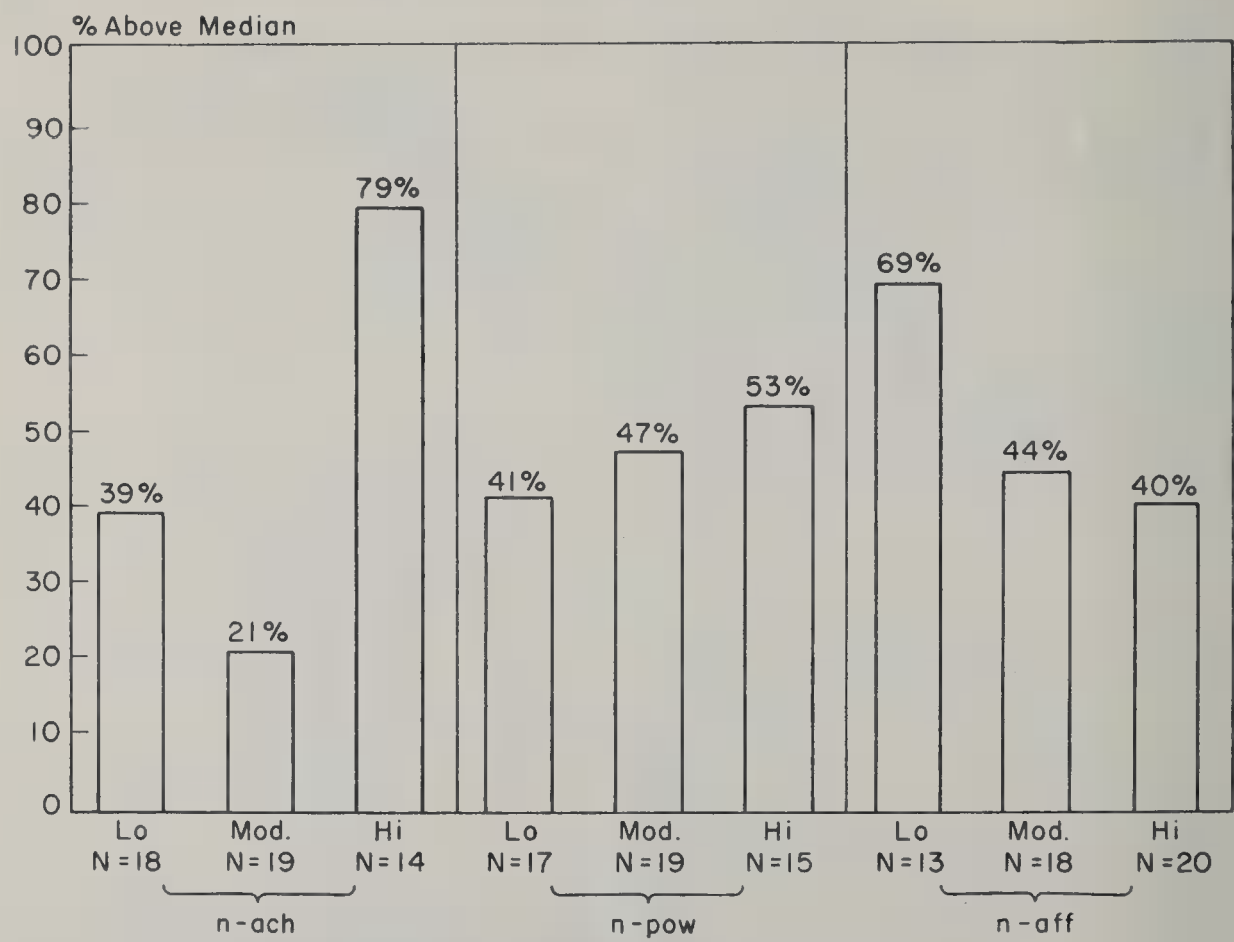


FIG. 1. Percentage of companies above median growth rate for total sample.

examining the data, it was noticed that, in addition to the very wide differences in company performance noted *between* high, moderate, and low n Ach groups, there existed substantial variations in company performance *within* each of these three groups. In other words, although the high n Ach group exhibited very high performance in comparison with the moderate and low n Ach groups, the range of performance scores *within* the high n Ach group was from .14 to 2.10. Similar within-group ranges were observed in the other two n Ach groupings.

An attempt was made, therefore, to determine whether these within-group variations could be attributed to variations in the strengths of the other two needs being investigated, n Pow and n Aff. The authors have further split the samples into high versus low performers (at the median performance score *within* each n Ach group) and compared levels of n Pow and n Aff within each of these new subgroups.

The following patterns emerge from the data summarized in Table 4. Within the low n Ach group, variations in performance are unaffected by variations in n Pow or n Aff. Within the moderate n Ach group, n Pow is identical for high versus low performers, while high performers within this group have a significantly higher n Aff. Finally, within the high n Ach group, n Aff is identical for high versus low performers, while high performers within this group have a significantly lower n Pow.

In summary, the *highest performing companies* in this sample were led by entrepreneurs who exhibited a high n Ach and a moderate n Pow. Those entrepreneurs who had a high n Ach coupled with a high n Pow performed less well than their high n Ach counterparts who exhibited only a moderate level of n Pow.⁶

⁶ When the authors use the phrases “moderate n Pow” or “high n Pow,” they are using as their reference point the distribution of scores observed in this study sample. Their specification, for example, of

TABLE 4

RELATIONSHIP BETWEEN PERFORMANCE AND *n* POW AND *n* AFF WITHIN HIGH, MODERATE, AND LOW *n* ACH GROUPS

| Need | High n Ach (≥ 9.0) | | Moderate n Ach ($4 \geq \times \leq 8$) | | Low n Ach (≤ 3.0) | |
|-------|----------------------------------|--------------------------------|---|------------------------------------|----------------------------------|--------------------------------|
| | Performance | | | | | |
| | Low ($\leq .59$) $N = 7$ | High ($> .59$) $N = 7$ | Low ($\leq .13$) $N = 9$ | High ($\geq .21$) $N = 10$ | Low ($\leq .26$) $N = 9$ | High ($> .26$) $N = 9$ |
| n Pow | 13.1(A) | 9.4(B) | 11.0(C) | 11.0(D) | 7.0(E) | 6.7(F) |
| n Aff | 2.0(G) | 2.0(H) | 2.2(I) | 5.5(J) | 4.9(K) | 4.0(L) |

Note.—Mann-Whitney *U* test (two-tailed): A versus B, $p < .08$; C versus D, $p < .40$; E versus F, $p < .50$; G versus H, $p < .50$; I versus J, $p < .02$; K versus L, $p < .25$.

Within the moderate *n* Ach group, higher performing companies were led by entrepreneurs who had a high *n* Aff.

DISCUSSION

The major hypothesis tested in this study predicted a positive relationship between an entrepreneur's level of *n* Ach and his company's performance. The authors' findings strongly support the conclusion that high *n* Ach is associated with high company performance, but the relationship between *n* Ach and performance is not linear across the entire range of *n* Ach scores. The relationship is markedly linear for the entrepreneurs whose *n* Ach is moderate to high. However, these entrepreneurs who scored low in *n* Ach were not significantly lower performers than those whose *n* Ach was moderate.

In an attempt to explain this nonlinearity it seems reasonable to assume that other needs or factors are influencing the entrepreneurial behavior of individuals who are not moderate to high in their level of *n* Ach. It is extremely likely that some threshold level of *n* Ach is necessary before one could assume that the strength of the need is significantly affecting the individual's behavior. In addition, it is obvious that the authors do not see *n* Ach as being the only (or for that matter the most important) factor that influences company

performance. They are arguing, however, that where the need exists in sufficient strength to influence entrepreneurial behavior significantly, company performance in general will improve.

A secondary aim in this study was to explore the question of whether a certain pattern or combination of needs was most often associated with high performance. In the introduction to this paper, it was suggested that *n* Pow and *n* Aff were needs whose behavioral manifestations were interpersonal in character. Satisfaction of these two needs, by definition, involves relationships with other people. *n* Ach, on the other hand, is much more individualistic in character. Satisfaction of one's *n* Ach, although often involving contact with other people, has behavioral manifestations which are qualitatively different in nature than either *n* Pow or *n* Aff.

The results of this study suggest that the combination of a high *n* Ach and a moderate *n* Pow characterizes the highest performing companies in the sample. In other words, a high (as opposed to moderate) level of *n* Pow appeared to counterbalance to some extent the positive benefits of a high level of *n* Ach.

One possible explanation for this finding lies in the relationship between *n* Pow and various styles of leadership. The lower an individual's *n* Pow, the more permissive or laissez-faire his style of leadership, the higher his *n* Pow, the more autocratic or authoritarian his style of leadership. The middle of the *n* Pow spectrum represents a mixed influence of the two extreme styles which is

high *n* Pow as being ≥ 13.0 was made prior to the analyses under discussion in this section. Consequently, classification of a mean *n* Pow of 13.1 as high and a mean *n* Pow of 9.4 as moderate is consistent with their a priori definitions.

best described as democratic.⁷ Prior research (Lippitt & White, 1958) has suggested that in certain situations the most effective leadership style is democratic and that performance of groups controlled in this manner is better than that of groups controlled by either of the other two styles.

Somewhat more difficult to explain is the finding concerning the positive differential effect on company performance, within the moderate n Ach group, of a high versus low n Aff level. It may be that for those individuals who have only a moderate level of n Ach, a high level of n Aff enables them to form close interpersonal relationships with their colleagues. In this way, the moderate n Ach individual may be able to acquire the assistance he needs from his colleagues, some of whom may well have a higher level of n Ach than he himself has.

⁷ The authors have assumed, of course, that high n Pow leaders are more likely to exercise an autocratic style of leadership and low n Pow leaders a laissez-faire style.

Interpretations in this area of need combinations must be viewed, at this point, as speculative and suggestive of further research. Analysis of the results of this study indicates that more complex relationships do have to be examined if a realistic view of performance determined by personality is to be gained. Future research should include replications of this study and the use of larger samples for the investigation of these hypotheses.

REFERENCES

- ATKINSON, J. W. *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.
- LIPPITT, R., & WHITE, R. K. An experimental study of leadership and group life. In *Readings in social psychology*. New York: Holt, Rinehart & Winston, 1958.
- MCCLELLAND, D. C. *The achieving society*. Princeton: Van Nostrand, 1961.
- SCHRAGE, H. The R&D entrepreneur: Profile of success. *Harvard Business Review*, 1965.

(Received January 22, 1968)

RATEE RELEVANCE IN PEER NOMINATIONS¹

FRANK T. PASSINI² AND WARREN T. NORMAN

University of Michigan

A new index, r_p , is proposed for the evaluation of ratee relevance in peer-nomination data. This index is shown to be a continuous linear function of the observed and random expected variance. The use of the r_p index for editing, weighting, diagnosing, and evaluating saliency is discussed.

A basic assumption that has been made by investigators using peer-nomination techniques is that the nominations or ratings they obtain veridically reflect the status of the ratees. This assumption was brought into serious question by Passini and Norman (1966). In the 1966 study Passini and Norman factor analyzed peer-nomination data that had been obtained from complete strangers. The factor structure that emerged from this analysis was highly similar to the factor structures that had emerged from previous analyses using the identical instrument with individuals who were well acquainted. As a tentative explanation of these rather surprising results the authors suggested the concept of shared implicit personality theories among raters. They said:

... if we accept the position that each rater brings to the situation an implicit personality theory which in certain aspects is similar to that of the other persons in the group and if observable features of the dress and manner of the participants are sufficient to provide an entree to one or more components of each of these common attribute clusters, then the interrater agreement and factorial structure obtained in the present study begins to seem a little less incredible [Passini & Norman, 1966, p. 48].

Granting the explanation offered by Passini and Norman, the problem that remains is to determine the degree of veridicality in any set of ratings of ratees. In 1966 Norman and Goldberg confirmed the fact that the same factor structure obtained from well-acquainted Ss could be obtained when the raters had absolutely no contact with the ratees. They used a Monte Carlo technique which simulated ratee-

rater independence while preserving the concept of raters having a shared conception of trait organization. The main thrust of Norman and Goldberg's effort, however, was the development of two criteria for the evaluation of the extent of ratee relevance in peer-nomination data.

These two measures were dubbed "score reliability" (r_s) and "rating reliability" (r_r). Score reliability is an index derived from a comparison of the expected random variances in a set of ratings and the obtained variance. Rating reliability is computed from (r_s) by reversing the generalized Spearman-Brown formula. Both reflect the degree of interrater agreement in the data. Norman and Goldberg (1966) applied these techniques to the Monte Carlo data in addition to four sets of empirical data. The obtained values were seen to reflect the length of acquaintanceship and intimacy among Ss used; that is, values of near zero were obtained for the Monte Carlo data (no acquaintanceship), while high values were obtained from the data generated by Peace Corps trainees.

The formula the authors gave for computing r_s is $r_s = 1 - \frac{V_r}{V_o}$ where V_r = random expected variance and V_o = observed variance. It can be seen readily that when there is perfect agreement among the raters as to a particular ratee's status the observed variance will be zero, and the defining expression of r_s becomes undefined.

Because of this discontinuity and the potential difficulty of interpreting a nonlinear function of observed variance, the present authors would like to suggest another index, r_p , as a replacement for r_s . The authors will attempt to show that this index is more directly interpretable as an index of interrater agree-

¹ This study was supported in part by Research Grant MH 07195 from the National Institute of Mental Health, United States Public Health Service, Warren T. Norman, project director.

² Requests for reprints should be sent to Frank T. Passini, who is now at the University of Colorado, Cragmor Road, Colorado Springs, Colorado 80907.

ment and in addition can readily be used to edit and weight data obtained from peer-nomination formats as described above. Also, the use of r_p as a diagnostic tool will be discussed briefly.

As Norman and Goldberg (1966) have pointed out, the estimate of ratee relevance is a function of the relationship between random variance and observed variance as long as independence of rating is insured. The proposed new index, r_p , is also a function of these two values. Thus, r_p is defined as:

$$r_p = 1 - \frac{V'_o}{V'_r}$$

where V'_o is the variance of a row (the nominations for a single ratee), and V'_r is the random (expected) variance of any single row. It can be seen that this index will vary from a maximum of 1.00 when there is complete agreement among the raters about a certain ratee to a minimum of $-.50$ when there is maximal disagreement among the raters for groups of size $3x + 1$ (see Figure 1).³

The interpretation of positive values of r_p offers no problems; however, the interpretation of negative values warrants a few remarks. Negative r_p values indicate that there is more than just chance disagreement among the raters about the status of a particular ratee on a scale. This disagreement could arise from at least three sources: (a) The scale has been so constructed that it is capable of diametrically opposed interpretations; (b) subsets of raters have taken the scale and fitted it into their own idiosyncratic antagonistic implicit personality theories; (c) the ratee behaves toward different subsets of raters in a diametrically opposed manner.

If an investigator is interested in how traits, as measured by scales, cluster, then he can increase the effectiveness of his clustering procedures by editing out ratees who have a negative r_p value. Any such editing, of course, will affect the generalizability of the analysis re-

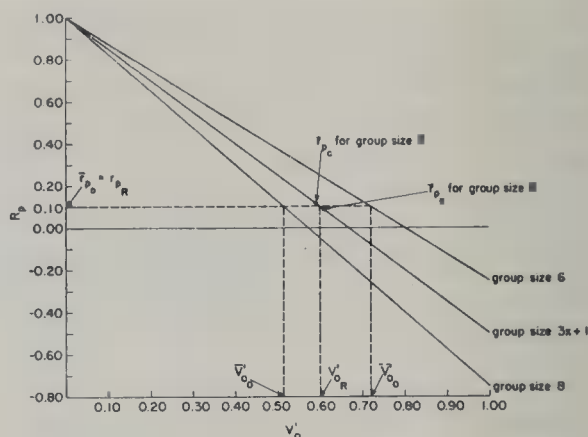


FIG. 1. Slope of the r_p function for groups of various sizes and slope correction.

sults and may substantially reduce the ns on which correlations may be computed for purposes of cluster analysis or factor analysis.

A not so apparent use of the r_p index is its application as a weighting factor to ratee scores. The rationale for this use is as follows: One would ordinarily be most certain of a ratee's status with respect to any given trait when all his peers agree as to his status on that trait. If one is to employ estimates of status on a trait, for whatever purpose, it might be reasonable to weight those estimates by a function of the certainty upon which they are based. Thus, r_p as an index of ratee relevance could be used as a certainty weighting factor.

The use of single r_p 's as tools for editing or weighting factors has been discussed thus far. Before pointing out another use of r_p , the statistic \bar{r}_p must be introduced: \bar{r}_p is the mean of the r_p 's for a given scale and as such is a measure of *average* scale relevance. This interpretation of \bar{r}_p follows from the fact that r_p is a linear function of V'_o (Figure 1) and acceptance of r_p as an index of the scale's relevance for the individual ratees.

As an example of the use of \bar{r}_p and r_p , suppose one has a scale that has a low \bar{r}_p value but a single high r_p value. Given a scale to measure subtle variable x which is related to deviant behavior y , a high r_p value coupled with a low \bar{r}_p value would be highly diagnostic for the individual. Conversely, if developing an instrument with new scales, a low \bar{r}_p with a few high r_p 's would be diagnostic for the scale. That is, intensive study of the individuals with high

³ It should be noted that under conditions of forced nomination as previously described, V'_r is a constant for groups of the same size. For all groups whose sizes are a multiple of $(3x + 1)$, the value of V'_r is a constant. For groups whose sizes are a multiple of $3x$ or $(3x + 2)$, the value of V'_r approaches this same constant as a function of group size (see Figure 1).

r_p 's could serve to highlight those characteristics to which the scale is mainly sensitive.

The above use of \bar{r}_p suggests an additional use for this index, that is, the evaluation of scale saliency across rating groups. There are, however, some problems in combining \bar{r}_p 's. First, it should be recognized that the range of \bar{r}_p will vary as a function of n . This variability of range, however, does nothing to the quality of scale values for groups of the same size. There is a further problem in combining \bar{r}_p 's from groups of different sizes in that the slope of the r_p function varies with group size. A correction factor based on the Pythagorean theorem may be applied to equate coefficients for different groups.

The intention of the correction factor is to equate the slopes of the r_p functions for varying size groups. Any slope could be selected as a standard, but for simplicity the slope of a group of size $3x + 1$ is selected as the reference (see Footnote 3). The correction formula is

$$\bar{r}_{p_o} = \bar{r}_{p_r} \frac{\sqrt{(V'_{or})^2 + (100 - r_{pr})^2}}{\sqrt{(\bar{V}'_{oo})^2 + (100 - \bar{r}_{p_o})^2}}$$

where \bar{r}_{p_o} is the corrected index, \bar{r}_{p_r} is the observed \bar{r}_p , V'_{or} is the reference observed variance, r_{pr} is the reference r_p , \bar{V}'_{oo} is the obtained \bar{V}'_o corresponding to \bar{r}_{p_o} . Note $\bar{r}_{p_o} = \bar{r}_{p_r}$. The

correction factor appearing in the formula is the ratio between square roots of the sum of the squared sides of two similar triangles. This correction factor will equate all observed \bar{r}_p 's to a common standard and therefore allow direct combining for purposes of comparison. For positive values of \bar{r}_p when group size exceeds 12, the correction would seem unnecessary due to the very small magnitude of change. Two examples of the correction are illustrated in Figure 1 for groups of six and eight with an observed \bar{r}_p of .10 (see Figure 1).

In view of the linearity of the r_p function and the utility of the r_p index for editing, weighting, diagnosing, and evaluating saliency, this index should be of considerable use in the development of new peer-nomination instruments. In addition, the r_p function should serve as a useful adjunct in the evaluation of data obtained from peer-nomination instruments in current use.

REFERENCES

- NORMAN, W. T., & GOLDBERG, L. R. Rater, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, 1966, 4, 681-691.
- PASSINI, F. T., & NORMAN, W. T. A universal conception of personality structure? *Journal of Personality and Social Psychology*, 1966, 4, 44-49.

(Received March 1, 1968)

SELF-ESTEEM AS A MODERATOR IN VOCATIONAL CHOICE: REPLICATIONS AND EXTENSIONS¹

ABRAHAM K. KORMAN²

New York University

The purpose of the research reported in this paper was to test in different types of vocational choice situations the hypothesis that self-esteem operates as a moderator on the vocational choice process in that high self-esteem (HSE) individuals are more likely to seek self-fulfillment than are low self-esteem (LSE) individuals. Four separate studies were all supportive of the proposition.

There are now several studies in research literature which support the hypothesis that the extent to which individuals choose careers which are need-fulfilling and those in which they believe they will be adequate is a positive function of the self-esteem of the individual (Korman, 1966, 1967a). Such differential choice patterns have been hypothesized to result from tendencies toward "balance" where individuals who perceive themselves as need-fulfilling and adequate (i.e., have HSE) choose vocational roles where they will have their needs fulfilled and will be adequate. On the other hand, situations of self-perceived need-fulfillment and adequacy are not "balanced" situations for those who have LSE; hence they do not serve as incentives for them.

Since the implications of this hypothesis, should it continue to be supported, have considerable importance for counseling processes as well as theoretical significance, it was felt that further testing of the proposition was desirable. A number of further studies were undertaken in order both to replicate these previous findings using different instruments and to extend them to different dimensions of the vocational choice process. It is the purpose of this paper to report these studies.

Study 1 consisted of testing the hypothesis that HSE individuals who enter a given occupation are more likely to describe themselves according to generally given stereo-

types of that occupation than both LSE people who enter that occupation and a random sample of those who enter different occupations. A second prediction was that there would be no difference between the latter two groups. In addition, it was felt that this hypothesis would hold for either specifically defined occupational choices (e.g., sales and accounting, as in our previous research) or grosser defined choice (e.g., business management in general). This latter aspect constitutes an attempt to extend previous findings.

The specific predictions which were made (in line with the above) were as follows: (a) HSE individuals in sales were most likely to describe themselves as being "sociable," "talkative," "aggressive," and having "initiative"; (b) HSE individuals in accounting were most likely to describe themselves as being "precise," "self-controlled," "organized," and "thorough"; (c) HSE individuals in general business were most likely to describe themselves as being "practical," "rational," and "responsible."

Study 2 also attempted to generalize the previous research to those whose occupational choice was more generalized in nature in that the interest here was in those whose occupational choice was "the world of business" rather than the more specific roles of sales, accounting, personnel, etc. The prediction was that HSE individuals who had chosen business as a career would be different from HSE individuals who had chosen something other than business as a career in the direction of having greater need for material security and less need for social service. On

¹ Studies 2, 3, and 4 in this paper were presented at the meeting of the American Psychological Association, Washington, D. C., 1967.

² Requests for reprints should be sent to the author, Department of Psychology, New York University, 21 Washington Place, New York, New York 10003.

the other hand, there should be no difference on these dependent variables between LSE individuals who had chosen business and LSE people who had not chosen business.

Study 3 consisted of a replication of Studies 1 and 2 in the area of "numerical abilities." It was predicted that individuals of HSE who had chosen numerically-oriented occupations (accounting and/or statistics) would see themselves as having higher numerical abilities than HSE individuals who had not chosen numerically-oriented occupations. On the other hand, such differences should not exist for those of LSE.

Study 4 proceeded from the assumption that the desire to engage in what is perceived to be ethical behavior is relatively widespread in nature, at least on a conscious self-descriptive basis. Hence, it was hypothesized that for individuals of HSE, the perceived ethicality of the behavior to be engaged in would be predictive of job choice, whereas such predictions would break down for those of LSE. More specifically, it was predicted that HSE individuals who had chosen business occupations would rate various business behaviors as being more ethical than HSE individuals who had not chosen business occupations. On the other hand, for LSE individuals, there would be no relationship between the judged ethicality of business behaviors and occupational choice.

METHOD

Sample

(a) The Ss for Study 1 consisted of male students at a far western state university. Of these, 22 had made career commitments to marketing and sales, 35 to accounting, 36 to general business, and 42 to some area unrelated to business. (b) The "business career" sample for Study 2 consisted of 65 male upper division business school majors at a different far western state university who indicated a specific commitment to enter the world of business, but who had not chosen a specific career such as sales, accounting, etc. The "nonbusiness career" sample consisted of 58 upper and lower division male students in a number of different major areas at two eastern universities. Most prominent of the major areas were education, the social and biological sciences, and social work. (c) The sample for Study 3 consisted of 67 lower division students at a private eastern university who had made a definite occupational choice. (d) The sample for Study 4 consisted of 53 males and 29 females. Since previous

work by the author has shown a sex difference in the ethical judgments used in this study, sexes were analyzed separately.

Measuring Instruments

(a) Self-esteem in all studies was measured by the Self-Assurance Scale of the Ghiselli Self-Description Inventory, with the cutoff in all cases for "high" and "low" self-esteem the 50th percentile on the nationwide norms. (b) The tendency to describe oneself according to certain adjectives in Study 1 was measured by the Gough Adjective Checklist (Gough, 1952). (c) The need for "material security" and "social service" in Study 2 were measured by the scales of the same name of the Crites Vocation Reaction Survey (Korman, 1966). (d) Self-perceived "numerical abilities" were measured by the Ability Assessment Questionnaire. This is an instrument of self-perceived abilities described by Korman (1967a). (e) "Occupational choice" was measured by questionnaire procedures found in previous research to have high reliability and concurrent validity (Korman, 1966). (f) Judgments of the "Ethicality of Business Behavior" in Study 4 were measured by having Ss rate, on a 4-point scale, the ethicality of 25 incidents which have actually occurred in the business world in recent years.

Procedures

The general procedure in all cases was to administer the questionnaires in regular class meetings or as part of introductory psychology research participation requirements. No systematic difference has been found between any of these procedures.

RESULTS

Study 1

Table 1 presents the results from this study, showing in all cases strong support for the hypothesis. The HSE individual does, in all cases, describe himself more as meeting the occupational image in the specific occupation than does the LSE individual in the occupation or a random sample of individuals who have made different occupational choices. Furthermore, this occurs no matter which occupation is referred to.

There was one possibly contaminating factor to these results, that, perhaps, the higher the self-esteem, the higher the frequency of words used to describe oneself. In other words, in a free response situation, such as the Gough Adjective Check List, where the person is asked to describe himself according to a set of adjectives with sheer number of adjectives chosen uncontrolled for, perhaps the only difference between HSE and LSE

TABLE 1

| ADJECTIVE SELF-DESCRIPTIONS ACCORDING TO OCCUPATIONAL CHOICE AND SELF-ESTEEM | | | |
|---|--|-----------------------|-----------------------|
| Item | High self- esteem | Low self- esteem | Random sample |
| Frequency of choosing "Practical," "Rational," and "Responsible" | | | |
| | 1. Business majors | 2. Business majors | 3. Nonbus majors |
| <i>M</i> | 2.73 | 1.90 | 2.12 |
| <i>SD</i> | .60 | .84 | .94 |
| <i>N</i> | 15 | 21 | 42 |
| | $t_1 \text{ \& } 3 = 2.35^{**}$ $t_2 \text{ \& } 3 = .92$ | | |
| Frequency of choosing "Precise," "Self- Con- trolled," "Organized," and "Thorough" | | | |
| | 1. Acct. majors | 2. Acct. majors | 3. Nonacct. majors |
| <i>M</i> | 2.35 | 1.33 | 1.60 |
| <i>SD</i> | 1.21 | 1.53 | 1.43 |
| <i>N</i> | 14 | 21 | 42 |
| | $t_1 \text{ \& } 3 = 1.79^*$ $t_2 \text{ \& } 3 = .69$ | | |
| Frequency of choosing "Initiative," "Aggres- sive," "Sociable," and "Talkative" | | | |
| | 1. Sales majors | 2. Sales majors | 3. Nonsales majors |
| <i>M</i> | 2.43 | 1.77 | 1.78 |
| <i>SD</i> | 1.34 | 1.29 | 1.13 |
| <i>N</i> | 9 | 13 | 42 |
| | $t_1 \text{ \& } 3 = 3.61^{**}$ $t_2 \text{ \& } 3 = .03$ | | |

Note.—All tests in this table are one-tailed tests.
* $p < .05$.
** $p < .01$.

people is in the total number of words chosen. A check for this, using frequency of choosing a random sample of 30 adjectives, indicated that this was not the case. The mean frequency of choice was exactly the same for the two self-esteem groups, carried out to one decimal place. Hence, this cannot explain the results.

Study 2

Table 2 presents the results of this investi- gation, with strong support once again being indicated for both hypotheses. The HSE busi- ness group is significantly higher than the HSE nonbusiness group on "material secu- rity" and significantly lower on "social service." For the LSE groups neither result occurs, with the means actually reversed for "material security."

Study 3

Table 3 presents the results for this study with strong support once again. HSE "quan- titative occupation" individuals see themselves as having greater numerical abilities than HSE "nonquantitative occupations," whereas the differences are not significant for the LSE groups.

TABLE 2
NEEDS FOR MATERIAL SECURITY AND SOCIAL SERVICE
ACCORDING TO SELF-ESTEEM AND
VOCATIONAL CHOICE

| Item | High self-esteem | Low self-esteem | High self-esteem | Low self-esteem |
|-------------------|---|-----------------|------------------|-----------------|
| Material security | | | | |
| | 1. Business | 2. Business | 3. Non-business | 4. Non-business |
| <i>M</i> | 7.59 | 6.08 | 6.07 | 6.61 |
| <i>SD</i> | 3.2 | 3.6 | 3.3 | 3.2 |
| <i>N</i> | 29 | 36 | 27 | 31 |
| | $t_1 \text{ \& } 3 = 1.77^*$ $t_2 \text{ \& } 4 = .65$ | | | |
| Social service | | | | |
| | 1. Business | 2. Business | 3. Non-business | 4. Non-business |
| <i>M</i> | 4.13 | 5.53 | 6.89 | 6.55 |
| <i>SD</i> | 2.9 | 3.6 | 4.4 | 4.2 |
| <i>N</i> | 29 | 36 | 27 | 31 |
| | $t_1 \text{ \& } 3 = 2.65^{**}$ $t_2 \text{ \& } 4 = 1.09$ | | | |

Note.—All tests in this table are one-tailed tests.
* $p < .05$.
** $p < .01$.

TABLE 3

SELF-PERCEIVED NUMERICAL ABILITIES ACCORDING TO SELF-ESTEEM AND VOCATIONAL CHOICE

| Item | 1. High self-esteem quantitative | 2. Low self-esteem quantitative | 3. High self-esteem nonquantitative | 4. Low self-esteem nonquantitative |
|-----------|----------------------------------|---------------------------------|-------------------------------------|------------------------------------|
| <i>M</i> | 13.00 | 10.56 | 8.69 | 9.52 |
| <i>SD</i> | 1.1 | 2.53 | 2.9 | 3.1 |
| <i>N</i> | 7 | 16 | 23 | 21 |
| | | $t_1 \text{ \& 3} = 3.99^*$ | | |
| | | $t_2 \text{ \& 4} = 1.11$ | | |

Note.—All tests in this table are one-tailed tests.

* $p < .01$.*Study 4*

The data for this study were analyzed by computing the mean judged ethicality of each of the 24 incidents for each of the four groups. All hypotheses were supported. Male HSE business ($n = 13$) rated the incidents as more ethical than male HSE nonbusiness ($n = 10$) (sign test, $p < .01$), whereas there was no difference between the male LSE business ($n = 17$) and nonbusiness ($n = 13$) groups (sign test, $p = .27$). Similarly, female HSE business ($n = 5$) were higher than nonbusiness ($n = 8$) (sign test, $p < .02$), whereas there were no differences for the female business ($n = 7$) and nonbusiness ($n = 9$) (sign test, $p = .50$).

DISCUSSION

Taking in context the results of the four studies reported here and the results reported in previous research, (Korman, 1966, 1967a), there is a highly consistent trend of evidence which argues that people differing in self-esteem choose occupations differently. Basically, the high self-esteem person seems to look at himself and say "I like what I see and I am going to give it its desires and needs," whereas the low self-esteem person seems to say, when looking at himself "I do not like what I see and I am not going to give it its desires and needs." While this may be a slight oversimplification, it seems to summarize, in essence, the kinds of results found over a wide variety of different instruments, different samples, and differing levels

of choice specificity. It is further strengthened by our related finding that even when the LSE individual is provided with fulfillment of his desires, it does not lead to satisfaction on his part (Korman, 1967b). (However, continued fulfillment of his needs *might* lead to a reevaluation of self, and thus change his determinants of satisfaction. There is little research on this.)

At least two further questions of interest occur here. The first is why this behavior occurs. A second is on what basis LSE people make vocational choices, if not on the basis of need-fulfillment. In terms of the first question, there are, of course, a variety of explanations ranging from childhood training patterns not to be contradictory in behavior to conceptions of a need for "social reality" comparable to that of physical reality. For both of these cases, situations of inconsistency would then be anxiety-provoking and, hence, to be avoided.

Turning to the second question, perhaps the LSE individual attempts to implement the value of an "ideal self" rather than an "actual self," a possibility which would generally be consistent with the notion of the LSE person as an individual who dislikes himself.³ In addition, it may be that such "ideal self" fulfillment is more determinate of his job satisfaction than his "actual self" fulfillment. A second possibility is that his behavior may be, at least partially, a function of social norms, that is, he may choose and be satisfied according to perceived social norms as to what is desirable and what is undesirable. However, since these possible explanations are not inconsistent with one another, since the relationship between self-esteem and persuasibility may be more complex than this (cf. Cox & Bauer, 1964), and since little research is available on either of them, such conjectures at this time are speculative only.

³ I am indebted to Mr. Jeffrey Greenhaus for this suggestion.

REFERENCES

- COX, O., & BAUER, R. Self-confidence and persuasibility in women. *Public Opinion Quarterly*, 1964, 28, 453-456.

- GOUGH, H. *The Adjective Checklist*. Palo Alto, Calif.: Consulting Psychologists Press, 1954.
- KORMAN, A. K. The self-esteem variable in vocational choice. *Journal of Applied Psychology*, 1966, **50**, 479-486.
- KORMAN, A. K. Self-esteem as a moderator of the relationship between self-perceived abilities and vocational choice. *Journal of Applied Psychology*, 1967, **51**, 65-67. (a)
- KORMAN, A. K. Relevance of personal need satisfaction for overall satisfaction as a function of self-esteem. *Journal of Applied Psychology*, 1967, **51**, 533-538. (b)

(Received April 3, 1968)

Manuscripts Accepted for Publication in the
Journal of Applied Psychology

- Processing Auditory Information: Interference from an Irrelevant Cue: J. Richard Simon* and A. M. Small, Jr.: University of Iowa, Iowa City, Iowa 52240.
- Impact of Employee Participation in the Development of Pay Incentive Plans: A Field Experiment: Edward E. Lawler, III,* and J. Richard Hackman: Department of Administrative Sciences, Yale University, 2 Hillhouse Avenue, New Haven, Connecticut 06520.
- Correlates of Employee Evaluations of Pay Increases: J. R. Hinrichs*: IBM Corporation, 112 East Post Road, White Plains, New York 10601.
- Prediction of Job Performance from Assessment Reports: Use of a Modified Q-Sort Technique to Expand Predictor and Criterion Variance: Garland Y. DeNelsky* and Michael G. McKee: 1020 North Quintana Street, Arlington, Virginia 22205.
- A Simplified Method for Increasing Mail Sorting Efficiency: Thomas Rex Long, Dell Lebo,* and Dorothy Ogden: Child Guidance Clinic, 1635 St. Paul Avenue, Jacksonville, Florida 32207.
- On the Use of Alternative Organisms: Douglas A. Bernstein* and Thomas M. Alloway: University of Illinois, Psychological Clinic, Children's Research Center Building, Champaign, Illinois 61820.
- Relevance of Rater-Ratee Acquaintance in the Validity and Reliability of Ratings: Norman E. Freeberg*: Developmental Research Division, Educational Testing Service, Princeton, New Jersey 08540.
- Use of Time-Lapse Photography in Flight Performance Evaluation: Robert N. Isley* and Paul W. Caro, Jr.: Human Resources Research Office, Division No. 6 (Aviation), P. O. Box 428, Fort Rucker, Alabama 36360.
- Consumer Attitudes toward Auto versus Public Transport Alternatives: Frank T. Paine, Allan N. Nash,* Stanley J. Hille, and G. Allen Brunner: Department of Business Administration, University of Maryland, College Park, Maryland 20742.
- Communication Length as an Index of Communicator Attitude: A Replication: Arthur N. Wiens, Russell H. Jackson, Thomas S. Manaugh, and Joseph D. Matarazzo*: Department of Medical Psychology, University of Oregon Medical School, 3181 S. W. Sam Jackson Park Road, Portland, Oregon 97201.

* Asterisk indicates author for whom address is supplied.

PREDICTIVE POWER OVER TEN YEARS OF MEASURED SOCIAL SERVICE AND SCIENTIFIC INTERESTS AMONG COLLEGE WOMEN¹

LENORE W. HARMON ■

University of Wisconsin—Milwaukee

The use of "usual occupation" instead of "current occupation" as a criterion was used to study the predictive validity of the Women's SVIB. One hundred sixty-nine women who scored A or B+ on the Social Worker scale and a contrasting group of 125 who scored A or B+ on the Laboratory Technician scale in 1953-1955 were located in 1966-1967 and asked about their vocational history and current vocational commitment. Thirty-nine percent of the former group and 36% of the latter group reported "usual occupations" appropriate to their SVIB scores; however, 44% and 40%, respectively, reported no "usual occupation." Among those reporting some career commitment, the predictive validity of the Women's SVIB was essentially equal to the validity of the men's form, but the SVIB was of no help in identifying which women would report career commitment.

The predictive validity of the Strong Vocational Interest Blank (SVIB) for women has never been explored. Studies by Strong (1955) and Campbell (1966a) have established the predictive validity of the SVIB for men by using eventual occupations as the criterion. Virtually all men are employed so it is relatively simple to determine whether they are employed in occupations predicted by earlier SVIBs.

It is not so easy to pick a criterion for the women's SVIB. Women enter and leave the labor force intermittently over their lifetimes. Careers are interrupted, temporarily or permanently, because of marriages and families. Only 35% of the married women in the United States are currently employed although the probability that a woman will be employed in a career some time after her

marriage is increasing (United States Department of Labor, 1966). While vocational planning is more important than ever for women, "current occupation" is not an appropriate criterion against which to validate vocational planning tools such as the Women's SVIB.

This study assessed the predictive validity of the Women's SVIB by the method used by Campbell (1966a) in which he followed up men who had all obtained high scores on one SVIB scale, but it differed from Campbell's in that the criterion used was "usual" rather than "current" occupation. The women were asked to state their usual occupation, whether or not they were currently employed or, if employed, whether or not they were currently employed in that field. Thus, for criterion purposes, the busy wife and mother who listed her usual occupation as "teacher" was classified as a teacher; the woman who was currently employed as her husband's accountant in a new business venture but listed her usual employment as "librarian" was classed as a librarian; the well trained nurse with 8 yr. of experience who now devotes herself to her home and family and answered "none" to the question, "What is your usual employment?" was classed as a housewife.

METHOD

The SVIB profiles of women entering the University of Minnesota during 1953-1955 were in-

¹ This paper was presented at the April 1968 convention of the American Personnel and Guidance Association in Detroit, Michigan.

² This study was begun at the University of Minnesota and continued after the author's departure. The Student Counseling Bureau, the Center for Interest Measurement Research, and the Director of Student Life Studies, Dr. Ralph Berdie, at the University of Minnesota generously continued to support the research. Deanna Berkenpas and Iffat Shah, who are research assistants at the Center for Interest Measurement Research, were invaluable in locating Ss and collecting data.

Requests for reprints should be sent to the author, 219 Mitchell Hall, University of Wisconsin, Milwaukee, Wisconsin 53201.

TABLE 1

RATE OF RESPONSE OF THE SW AND LT GROUPS

| Item | SW group | | LT group | |
|-----------------------------------|----------|-----|----------|-----|
| | N | % | N | % |
| Completed questionnaires received | 169 | 55 | 124 | 51 |
| No response | 76 | 25 | 75 | 31 |
| Never located | 64 | 20 | 43 | 18 |
| Total | 309 | 100 | 242 | 100 |

spected to locate those who had high scores (A or B+) on the Social Worker (SW) scale. This scale was chosen because Layton (1958) found that 25% of one class of University of Minnesota freshmen obtained high scores on it, and because social work is a field women can enter relatively easily. This choice insured both a reasonable number of cases to work with and a fair trial for the scale.

For contrast, another group with high scores (A or B+) on the Laboratory Technician (LT) scale were also identified. This scale was selected because the Social Worker group scored lowest on it; the correlation between the two scales is $-.65$ (Campbell, 1966b). No Ss overlapped between the SW & LT groups.

The SW group originally included 309 women, and the LT group 242. Locating the women was difficult because most of them had married and moved from the addresses at which they had lived as students. Questionnaires were systematically mailed to all the former residences listed in university records for each woman in the SW group in the spring of 1966 and to women in the LT group in the spring of 1967. After one follow-up letter, the returns were as

reported in Table 1. The Never Located group includes women for whom no address at which mail was accepted was ever found. The No Response group includes women for whom mail was accepted at one of the addresses available; however, there was no assurance that mail was actually forwarded since there was no response from these addresses. These returns are lower than Campbell's 77% in the study of men with high scores on Life Insurance Salesmen, but locating women is more difficult because they change their names.

RESULTS

The mean SVIB profiles for the SW and LT groups are in Figure 1. For the SW group, the Speech Therapist and Music Performer scales were nearly as high as the Social Worker scale, the Medical and Physical Science groups were rejected, and the premarital pattern of high scores on the Housewife, Office Worker, Steno-Secretary scales (Layton, 1958) was not evident. For the LT group, scores on the Physical Therapist and Nurse scales were actually higher than scores on the Laboratory Technician scale; all high scores were in Health Service and Medical Science occupations; the Music, Verbal-Linguistic, Social Service, and Sales occupations were rejected, and the premarital pattern was not evident in this group either.

For each group the questionnaire responses regarding major in college, usual occupation, and current occupation, were classified intuitively into the following categories for the

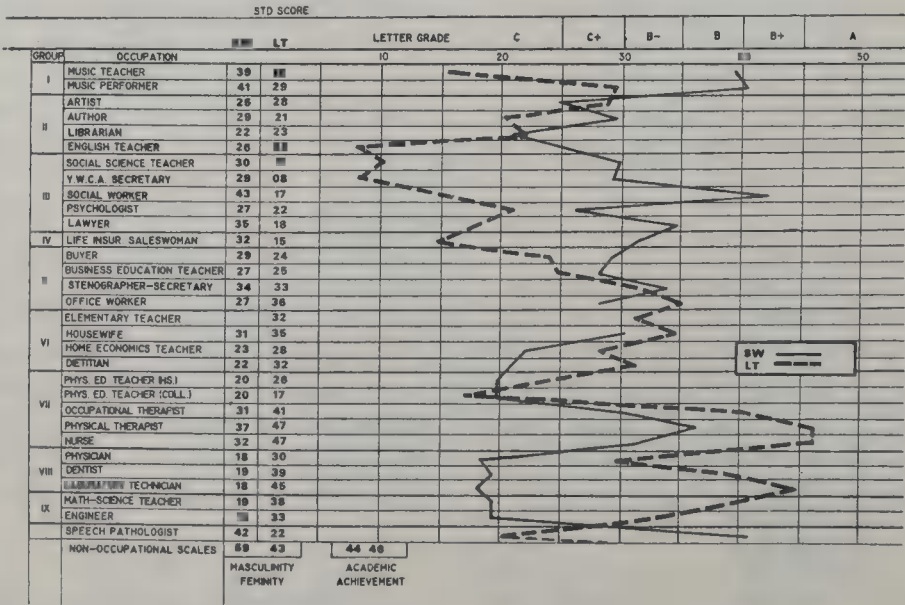


FIG. 1. Mean SVIB profiles of the SW and LT groups.

TABLE 2
COLLEGE MAJORS OF THE SOCIAL WORK GROUP

| Major | Career | | | Homemaker | | | Total | | |
|----------------------|----------|-----|--------------|-----------|-----|--------------|----------|-----|--------------|
| | <i>N</i> | % | <i>Cum %</i> | <i>N</i> | % | <i>Cum %</i> | <i>N</i> | % | <i>Cum %</i> |
| Social work | 18 | 19 | 19 | 12 | 16 | 16 | 30 | 18 | 18 |
| Teaching | 27 | 29 | 48 | 20 | 27 | 43 | 47 | 28 | 46 |
| Medical service | 17 | 18 | 66 | 2 | 3 | 46 | 19 | 11 | 57 |
| Other social science | 7 | 7 | 73 | 4 | 5 | 51 | 11 | 6 | 63 |
| Other-unrelated | 16 | 17 | 90 | 13 | 17 | 68 | 29 | 17 | 80 |
| None | 9 | 10 | 100 | 24 | 32 | 100 | 33 | 20 | 100 |
| Total | 94 | 100 | | 75 | 100 | | 169 | 100 | |

SW group: (a) Social Work, (b) Teaching, (c) Medical Service, (d) Other Social Service, (e) Unrelated Professional, and (f) Clerical and Skilled. For the LT group, the comparable categories were: (a) Medical Technology, (b) Other Medical, (c) Other Scientific and Mathematical, (d) Unrelated Professional, and (e) Clerical and Skilled.

Categories *a-d* for the SW group included occupations which are compatible with high scores on the Social Worker scale. These scores reflect an interest in direct, helping relationships such as the practice of social work, teaching, nursing, and other occupations like speech therapy and clinical psychology.

Similarly, Categories *a-c* for the LT group included occupations which are compatible with high scores on the Laboratory Technician scale. These scores reflect scientific and mathematical interests and preferences for indirect helping relationships such as the practice of medical technology, X-ray technology, and pathology.

Tables 2 and 3 show the college majors for the SW and LT groups separated into Career and Homemaker subgroups. The Career subgroups included those who listed a usual occupation, even if they were not currently employed, or were currently employed in some other job. The Homemaker subgroups included those who listed no usual occupation even if they were currently employed.

If one is willing to grant that a college major in any of the first four categories in Table 2 is consistent with a high Social Worker score, 63% of the SW group were enrolled in consistent college majors; the percentage is higher (73%) in the Career subgroup. Analogously, in Table 3, 56% of the LT group were enrolled in college majors consistent with their high scores on the LT scale, 63% of those in the Career subgroup.

Tables 4 and 5 show the usual and current occupations of the SW and LT groups. Nine percent of each group was perfectly predicted, that is, became either Social Workers or Laboratory Technicians.

TABLE 3
COLLEGE MAJORS OF THE LABORATORY TECHNICIAN GROUP

| Major | Career | | | Homemaker | | | Total | | |
|---------------------|----------|-----|--------------|-----------|-----|--------------|----------|-----|--------------|
| | <i>N</i> | % | <i>Cum %</i> | <i>N</i> | % | <i>Cum %</i> | <i>N</i> | % | <i>Cum %</i> |
| Med tech | 12 | 16 | 16 | 4 | 8 | 8 | 16 | 13 | 13 |
| Other medical | 26 | 35 | 51 | 11 | 22 | 30 | 37 | 30 | 43 |
| Other science, math | 9 | 12 | 63 | 7 | 14 | 44 | 16 | 13 | 56 |
| Other-unrelated | 25 | 33 | 96 | 15 | 30 | 74 | 40 | 32 | 88 |
| None | 3 | 4 | 100 | 13 | 26 | 100 | 16 | 13 | 101 |
| Total | 75 | 100 | | 50 | 100 | | 125 | 100 | |

TABLE 4
OCCUPATIONS OF THE SW GROUP

| Job category | Career | | | | | | Homemaker | | | Total | | | | | |
|------------------------|-------------|----|-------|-----------|----|-------|-------------|----|-------|-------------|----|-------|-----------|----|-------|
| | Current job | | | Usual job | | | Current job | | | Current job | | | Usual job | | |
| | N | % | Cum % | N | % | Cum % | N | % | Cum % | N | % | Cum % | N | % | Cum % |
| Social work | 6 | 6 | 6 | 16 | 17 | 17 | — | — | 0 | 6 | 4 | 4 | 16 | 9 | 9 |
| Teaching | 21 | 22 | 28 | 30 | 32 | 49 | 1 | 1 | 1 | 22 | 13 | 17 | 30 | 18 | 27 |
| Medical service | 11 | 12 | 40 | 17 | 18 | 67 | 1 | 1 | 2 | 12 | 7 | 24 | 17 | 10 | 37 |
| Other social service | 2 | 2 | 42 | 3 | 3 | 70 | — | — | 2 | 2 | 1 | 25 | 3 | 2 | 39 |
| Unrelated professional | 13 | 14 | 56 | 16 | 17 | 87 | 2 | 3 | 5 | 15 | 9 | 34 | 16 | 9 | 48 |
| Clerical, skilled | 9 | 10 | 66 | 12 | 13 | 100 | 2 | 3 | 8 | 11 | 6 | 40 | 12 | 7 | 55 |
| Unemployed | 32 | 34 | 100 | — | — | — | 69 | 92 | 100 | 101 | 60 | 100 | 75 | 44 | 99 |
| Total | 94 | | | 94 | | | 75 | | | 169 | | | 169 | | |

Note.—Percentages may total 99 or 101 due to rounding.

When the expanded criterion of “usual occupation” was used and all social service or scientific occupations were regarded as appropriate, the percentages increased to 39 and 36; among those with career commitments, the percentages were yet higher, 70 and 61.

The sets of categories for the SW and LT groups overlap. For instance, medical occupations appear in both because people can enter medical occupations out of social service interests, scientific interests, or some combination of the two.

Table 6 illustrates that there is a qualitative as well as a quantitative difference between

the SW and LT groups in the medical occupations which they list as their usual occupations. All of the members of the SW group who listed medical occupations chose occupations involving direct patient contact in which human relationships are of primary importance. Over half the LT group who listed medical occupations chose occupations in which patient contact is minimal or indirect, and scientific procedure is more important than human relationship. Thus the overlap between the medical occupations accepted as compatible with social service and scientific interests is not as great as it might seem; neither is it nonexistent.

TABLE 5
OCCUPATIONS OF THE LT GROUP

| Job category | Career | | | | | | Homemaker | | | Total | | | | | |
|-----------------------|-------------|----|-------|-----------|----|-------|-------------|----|-------|-------------|----|-------|-----------|----|-------|
| | Current job | | | Usual job | | | Current job | | | Current job | | | Usual job | | |
| | N | % | Cum % | N | % | Cum % | N | % | Cum % | N | % | Cum % | N | % | Cum % |
| Med tech | 7 | 9 | 9 | 11 | 15 | 15 | — | — | 0 | 7 | 6 | 6 | 11 | 9 | 9 |
| Other medical | 11 | 15 | 24 | 26 | 35 | 50 | — | — | 0 | 11 | 9 | 15 | 26 | 21 | 30 |
| Other scientific—math | 4 | 5 | 29 | 8 | 11 | 61 | — | — | 0 | 4 | 3 | 18 | 8 | 6 | 36 |
| Other professional | 15 | 20 | 49 | 24 | 32 | 93 | 5 | 10 | 10 | 20 | 16 | 34 | 24 | 19 | 55 |
| Clerical, skilled | 5 | 7 | 56 | 6 | 8 | 101 | 4 | 8 | 18 | 9 | 7 | 41 | 6 | 5 | 60 |
| Unemployed | 33 | 44 | 100 | — | — | — | 41 | 82 | 100 | 74 | 59 | 100 | 50 | 40 | 100 |
| Total | 75 | | | 75 | | | 50 | | | 125 | | | 125 | | |

Note.—Percentages may total 99 or 101 due to rounding.

If a way could be found to differentiate the women who would eventually develop career commitments from those who would retire permanently from employment, the predictive accuracy of specific scales could be increased. That is, if a counselor could identify which 60% of the women who get high scores on the Laboratory Technician scale when they enter college would develop career commitments, he could predict that about 15% of

| Medical occupation | SW group | | LT group | |
|---------------------------|----------|------------------|----------|------------------|
| | <i>N</i> | % of total group | <i>N</i> | % of total group |
| Nurse | 12 | 7 | 10 | 8 |
| Med technologist | — | — | 11 | 9 |
| X-ray technologist | — | — | 4 | 3 |
| Physical therapist | — | — | 4 | 3 |
| Occupational therapist | 4 | 2 | 4 | 3 |
| Pathologist (M.D.) | — | — | 2 | 2 |
| Veterinarian | — | — | 1 | 1 |
| Medical records librarian | — | — | 1 | 1 |
| Dental hygienist | 1 | 1 | — | — |
| Total | 17 | 10 | 37 | 30 |

them would choose Medical Technology and 61% of them would choose scientific occupations. These levels of accuracy approach the ones found for scales of the Men's SVIB (Campbell, 1966a; Strong, 1955).

Figure 2 shows that the mean SVIB profiles for the Career and Homemaker groups were quite similar. The Housewife, Academic Achievement, and Masculinity-Femininity scales which might be expected to differentiate the subgroups did not do so. An earlier study



(Harmon, 1967) suggested that the SVIB Housewife scale was not predictive of women's career patterns. Comparable data for the LT subgroups also shows a striking similarity between the Career and Homemaker groups. There is no way of differentiating between the Career and Homemaker groups on the basis of SVIB profile.

CONCLUSIONS

"Usual occupation" is a fair criterion for studying the predictive validity of vocational interest tests for women. It avoids the problems of using "current occupation" or "any experience in the occupation" as criteria. For women who are committed to careers, that is, those who claim some "usual occupation," the Social Worker and Laboratory Technician scales of SVIBs taken 10-14 yr. earlier accurately predicted commitment to social service or science occupations 70 and 61% of the time, respectively. For the 40-45% of women who are homemakers with no career commitment, the Social Worker and Laboratory Technician scales of SVIBs taken 10-14 yr. earlier were, by definition, not predictive. Their SVIB profiles did not differ much from those of women with career commitments.

It would be interesting to determine what kind of women develop career commitments and which do not. This should be done but, from a practical point of view, it is more im-

portant to determine the differences between occupational interests because the women who utilize testing to help make decisions are usually choosing among vocational plans. The woman who prepares for a career may decide to follow it or not. The woman who does not prepare has no choice. The counselor who uses the Women's SVIB as though each high score was predictively accurate for all women rather than just for those who become career women will leave more doors open for future choices by his client.

REFERENCES

- CAMPBELL, D. P. Occupations ten years later of high school seniors with high scores on the SVIB life insurance salesman scale. *Journal of Applied Psychology*, 1966, **50**, 369-372. (a)
- CAMPBELL, D. P. Strong Vocational Interest Blank manual. Stanford: Stanford University Press, 1966.
- HARMON, L. W. Women's working patterns related to the SVIB Housewife and "Own" Occupational Scores. *Journal of Counseling Psychology*, 1967, **14**, 299-301.
- LAYTON, W. L. *Counseling use of the Strong Vocational Interest Blank*. Minnesota Studies in Personnel Work, No. 8. Minneapolis: University of Minnesota Press, 1958.
- STRONG, E. K., JR. *Vocational interests 18 years after college*. Minneapolis: University of Minnesota Press, 1955.
- UNITED STATES DEPARTMENT OF LABOR. *1965 handbook on women workers*. Women's Bureau Bulletin No. 290. Washington, D. C.: United States Government Printing Office, 1966.

(Received April 8, 1968)

COGNITIVE, NONCOGNITIVE, AND ENVIRONMENTAL CORRELATES OF MECHANICAL INGENUITY¹

W. A. OWENS²

University of Georgia

This is a 1964 follow-up study of 1500 engineering students originally administered four tests of mechanical creativity during 1955. Analyses were devoted to the criterion validity of the original battery, the job environment and life history correlates of outstanding performance, and the possibility of a Type of Person \times Optimal Environment interaction. Results suggest (a) a validity coefficient of .40 for the tests of 1955; (b) factors of academic underachievement and research orientation of supervision as important personal and environmental correlates of performance; and (c) no Type of Person \times Optimal Environment interaction.

The construction of a special battery of tests designed to discriminate creative from noncreative or development engineers and the subsequent validation of these tests on an independent sample of 304 industrially employed engineers was reported by Owens, Schumacher, and Clark in 1957. The final form of the battery which survived cross-validation consisted of the *Personal Inventory* (PI), a quasi forced-choice inventory dealing with interests, attitudes, opinions, personal characteristics, and experiences; the *Personal History Form* (PHF), a single sheet dealing with personal background; the *Application of Mechanisms Test* (AMT); solutions to the *Power Source Apparatus Test* judged to be workable (PSA-W), and total number of solutions to the PSA (PSA-T). Conclusive demonstration of the predictive efficiency of the battery, however, required a longitudinal design which would permit the accumulation of evidence of creative performance over a number of years. With this purpose in mind, the final battery was administered in 1955 to over 1500 juniors and seniors in the mechanically-related branches of engineering at 25 colleges and universities. The security of the test scores was maintained while criterion information was accumulating and an account of the relationship between the predictions of

1955 and the actualities of 1964 follows. An attempt has also been made to recognize some of the moderating influences involved.

PURPOSES AND HYPOTHESES

Specifically, attention has been directed to three questions: (a) What is the evidence regarding the predictive validity of the 1955 measures of creativity in machine design? (b) What are the personal (nongognitive) and environmental characteristics which have facilitated or inhibited the expression of this creativity in the meantime? and (c) If we identify types of persons and types of environments, do they interact; that is, is it true that one type of environment is optimal for persons of Type A and another for persons of Type B? Attendant upon, and congruent with, the preceding purposes, the following hypotheses were formulated: (a) The present cognitive tests are better predictors of creativity in machine design than a common mental ability (or scholastic aptitude) test; this is, at least in part, because the one places premium upon a different cognitive pattern or style than the other (Guilford, 1959). (b) The creative individual is cognitively complex and can integrate more inputs than his less creative fellows; thus, tests which restrict or structure more, and which imply more inputs (PSA), will be superior to those which restrict or structure less (AMT). (c) Accepting the phenomenal nature and complex determination of creativity, prediction will be enhanced by appraising not only cognitive characteristics, but nongognitive and

¹ The author wishes to recognize the substantial contributions of Michael Brodie, Maureen Kallick, Stephen P. Klein, Richard Klimoski, Robert B. Means, and Mark Van Slyke.

² Requests for reprints should be sent to the author, Department of Psychology, University of Georgia, Athens, Georgia 30601.

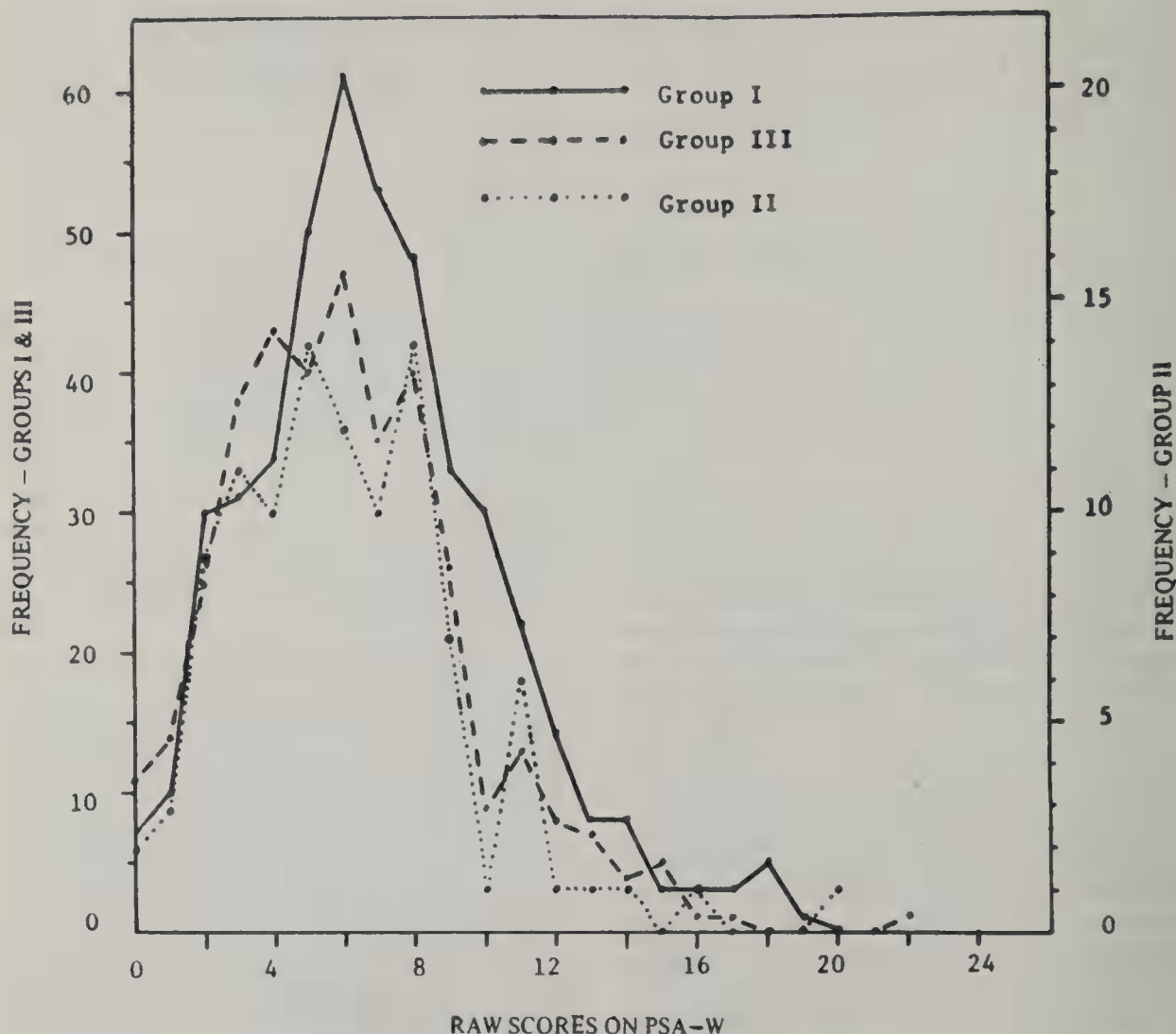


FIG. 1. A comparison of the shapes of distributions on the PSA-W test for Groups 1, 2, and 3.

environmental characteristics as well. (d) Since creative persons do not all appear to respond to the same contexts or to function in the same way, an interaction between personal and environmental determiners of creativity is postulated. Thus Environment X may be optimal for persons of Type A and Environment Y for persons of Type B.

METHODS

In overview the methods adapted to the above purposes are as follows: (a) Correlational analyses were utilized to answer Question 1 regarding the predictive validity of the 1955 battery. (b) To answer Question 2, inventories of personal (non-cognitive) and environmental characteristics were completed by Ss in 1964; the items of each were factored, and factor scores were then correlated with the creativity criterion. (c) Question 3 was attacked by successively subgrouping Ss; first

on the basis of the profile similarities of their job environment factor scores, and then on the basis of the profile similarities of their personal characteristics factor scores. The two classes of subsets then became the two criteria of classification for an analysis of variance in which the interaction would represent Type of Person \times Type of Environment.

Subjects

The potential pool of Ss for the follow-up of 1964 consisted of 1537 students of 1955 who were, typically, originally tested as beginning upperclassmen in the mechanically-related branches of engineering. The vast majority were in civil, aeronautical, or mechanical curricula, and they were from all parts of the country and a wide variety of institutions. Alumni offices of the schools in question were able to supply 1268 addresses. Of these, 109 proved to be insufficient or incorrect and no forwarding address could be obtained. Thus, 1159 Ss were actually contacted and 938 (81%) ultimately replied. Since seven returns were unusable, 931 Ss

TABLE 1

CRITERION ELEMENTS: THEIR MEANS, STANDARD DEVIATIONS, AND RELATIVE IMPORTANCE WEIGHTS

| LHQ item | Group 1 | | |
|--|---------|-------|-----|
| | M | SD | Wt. |
| 104. a. Improving products or processes | 5.28 | 11.44 | 1 |
| 104. b. Developing products or processes | 3.57 | 9.29 | 1 |
| 114. Papers presented at professional meetings | 1.27 | 0.68 | 2 |
| 121. Papers published in professional journals | 0.71 | 2.07 | 2 |
| 126. In S's own name: | | | |
| a. Patents held | 0.06 | 0.32 | 3 |
| b. Patents pending | 0.21 | 0.81 | 3 |
| c. Patent disclosures | 0.42 | 1.50 | 4 |
| 127. With contribution by S: | | | |
| a. Patents held | 0.12 | 0.68 | 1 |
| b. Patents pending | 0.11 | 0.62 | 1 |
| c. Patent disclosures | 0.14 | 0.64 | 2 |

Criterion

The 10 items of criterion information collected appear in Table 1 along with their means, standard deviations, and relative importance weights. These weights were assigned by five members of the engineering faculty at Purdue all of whom had had both academic and industrial experience relevant to the design and development of products or processes. In the context of their importance to these activities, the weights were assigned to each criterion element by each judge in accordance with the Kelly "bids system" (see Toops, 1944). The criterion score of a given S, then, was his standard score on the given criterion element, multiplied by the appropriate importance weight and summed across the 10 elements.

Statistical Treatment

Treatment of the data in the context of the three purposes originally stated would probably be most intelligible. Thus the first purpose was to answer certain questions regarding the predictive validity of the creativity battery of 1955. As noted in the methods overview, these questions were answered in terms of correlational analyses of conventional character.

The second purpose was to identify some of the personal (nongognitive) and environmental characteristics which have facilitated or inhibited the subsequent expression of the creativity measured by the battery of 1955. This purpose was served by factoring independently the items of the JES and the LHQ and by relating factor scores on the factors obtained to the composite criterion of creativity. In the case of the JES most of the items are answered as applying or not applying (1 or 0) to the given S's working situation. However, since some permit multiple responses, a total of 109 options were available for analysis. No continuum-type items were included, and 10 binary items were eliminated because they yielded response frequencies below 10%, thereby enhancing the risk of obtaining difficulty factors. Thus, 99 item-options (hereafter "items") ultimately entered the factor analysis. Twenty-one factors were initially extracted from the matrix according to the method of principal components. A decision to rotate four of these was based on a plot of the latent roots. Following an orthogonal rotation, all of the items which did not load at least .20 on one of the rotated factors were eliminated as were most items which loaded nearly equally across several factors. This procedure was then repeated until nearly half of the original items had been dropped. The technique was adopted for two reasons: (a) It was felt that it would sharpen interpretations regarding the differential criterion relevance of various environmental influence factors; (b) It was planned to fulfill the third purpose of this study by classifying Ss on their factor score profiles in a fashion requiring the independence of the basic dimensions. In anticipation of this next step, standardized factor scores were computed for each individual on each of the ultimate job environment factors.

were involved in this study. These were assigned to one of the following three categories.

Group 1 was composed of 457 engineers employed in research and development (R & D). Group 2 consisted of 104 former R & D engineers promoted into engineering management (EM). Group 3 was composed of 370 Ss in two loosely defined subgroups: first, those not in engineering per se, but in an engineering-related occupation, such as teaching engineering or sales engineering; second, those in an unrelated area, such as medicine, law, or the ministry.

Figure 1 essentially involves a comparison of the shapes of the test score distributions of the three groups of Ss on the single most valid measure. However, in speaking of the predictive purposes of the study, it must be clear that Ss, by groups, have not had an equal opportunity to accumulate criterion evidence of their creativity. Attention has, therefore, been centered on Group 1 (R & D).

Measuring Instruments

In the follow-up of 1964, two inventories and a covering letter were sent to each S. Many investigators have reported relevant and favorable experience with the biographical information blank, or BIB, as a predictor (Owens & Henry, 1966). Accordingly, the first inventory was a so-called *Life History Questionnaire* (LHQ) of 181 items dealing with Ss' demographic characteristics and experiential background plus 10 criterion items (Klein, 1965). Pursuant to a suggestion by Taylor (1964), the second was a *Job Environment Survey* (JES) of 80 items covering the "research climate" in which S worked (Kallick, 1964).

An identical procedure was utilized with the items of the LHQ, with the following two exceptions. (a) The LHQ contains both continuum-type and binary items; however, the latter have several options under the same stem from which only one is to be selected. The options are, thus, not invariably independent and only one from a common stem could be permitted to enter the factor analysis. (b) Since mixing binary and continuum-type items leads to ambiguities, it was decided to factor the binary items first and to include binary factor scores, along with scores on continuum-type items as entries in the matrix to be factored. When the final set of three personal history factors had been identified, standardized factor scores were computed for each *S* on each factor. It was then a simple matter to fulfill the second purpose regarding the relevance of personal and environmental factors to creativity by correlating scores on each of the two sets of factors with the composite criterion.

Finally, the third purpose had reference to testing the possibility of a Type of Person \times Type of Environment interaction. To test it required that *Ss* be subgrouped, successively, on each set of factors and assigned to appropriate cells in an analysis of variance design with two criteria of classification. On the job environment factors, for example, this was accomplished as follows. Each *S's* factor scores were regarded as comprising a profile. The similarity of each profile to every other profile was expressed in terms of the D^2 statistic of Cronbach and Gleser (1953). Subgrouping was then accomplished through application of the hierarchical procedure of Ward and Hook (1963) to the obtained matrix of D^2 values. The technique essentially compares each profile with every other, combines the most similar pair, treats them as a unit, and repeats the process. The criterion is the minimization of within-groups variance and the machine output indicates the increase in this variance which accompanies each reduction in the number of subgroups. Since the operation is sequential and the decision regarding the ultimate number of groups judgmental, no claim is made that the overall solution is optimal.

For these reasons a special evaluation program was written (Brodie, 1966) to follow the Ward-Hook. Via the D^2 this program compares each *S's* profile with the mean profile of each potential subgroup and permits several important outcomes: (a) The *Ss* may be reassigned to a new subgroup if this is indicated; (b) the *Ss* who equally resemble two or more subgroups within specified limits may be placed in a residual group; and (c) errors made at several grouping levels may be used as one objective criterion for determining the optimum number of groups. The evaluation program was applied to the present data and the results were used to serve the three purposes indicated.

When *Ss* had been subgrouped in the indicated fashion on the basis of the job environment factors, the same procedure was repeated employing the life history factors. The former subgroups were then regarded as representing levels of a "job climate"

factor and the latter as levels of a type of person factor in an unequal cell-size, weighted means, analysis of variance design. Cell entries were composite criterion scores for a given type of *S* exposed to a given climate. An interaction between these two criteria of classification would then be identifiable and would speak directly to the third purpose of the study (see Means, 1966).

RESULTS

In attempting to evaluate the predictive validity of the creativity battery of 1955, two preliminary issues arise. First, before examining the relationship of predictors to criteria within a selected group (R & D), it seems appropriate to inquire as to the distributions of the three groups on the single best predictor, PSA-W. Since the distributions concerned are skewed, their means and variances do not define them adequately. The complete scatter plots have been included as Figure 1. It should be noted that two different scales are employed on the *y* axis in order to make the areas under the curves more nearly comparable. Given this, there is an apparent tendency for Group 3 to show the lowest scores and Group 1 the highest, as expected. There may also be some tendency for Group 2 to score more like 3 than 1, a result which might or might not have been anticipated.

Second, before examining in greater detail the issue of how well the creativity battery predicts, it seems a prerequisite that it predict with some uniqueness. The question most commonly raised concerns the independence of what is measured by general mental ability tests from what is measured by creativity tests. In the present instance it was found possible to obtain scores on the American Council on Education Psychological Examination (ACE) for 167 of the 457 R & D engineers. The test had been administered to them as college freshmen some 2 yr. prior to the time when they completed the creativity battery. The top row of Table 2 contains the correlations of the ACE with the cognitive predictors of the creativity battery and with the composite criterion. It will be noted that none of the former is of substantial magnitude and that the latter is near zero. On the other hand, those scores derived from the special battery are significantly, if modestly, cor-

related with the composite criterion. The result should not be overinterpreted, but it does suggest reasonable independence of creativity from general mental ability within the present sample.

To evaluate the predictive power of the creativity battery over time, reference is made to Column 5 of Table 2 which contains minimum estimates of the relationships involved. It will be observed that only the PSA test and the PHF are effective predictors. Both the AMT, which looked promising in concurrent validation, and the PI, which had appeared to be making a marginal contribution, failed to correlate significantly with the criterion in the more demanding predictive context. At any event, the estimates of Column 5 can, realistically, be increased if consideration is given to three qualifying influences.

First, unambiguous criterion data are available only for the R & D group. However, if the battery were administered as an employment test immediately following graduation, the range of scores would be wider because both Groups 1 (R & D) and 2 (EM) would be included. If it were administered as engineering students became upperclassmen, to serve as a sectioning device, the range would be still wider because Group 3 would also be included. Even in the first case, the zero-order criterion correlations of Table 2 would be .02 to .03 points higher in the more variable sample (see Thorndike, 1949, pp. 169-173).

TABLE 2

PREDICTIVE VALIDITIES AND INTERCORRELATIONS OF THE ACE AND SEVERAL SPECIALLY DEVISED PREDICTORS

| Predictor | 2 | 3 | 4 | 5 |
|--------------------------------|-----|------|------|------|
| ACE ($N = 167$) | .07 | .32* | .20* | .01 |
| Application of Mechanisms Test | | .14* | .40* | .06 |
| Power Source Apparatus (PSA-W) | | | .63* | .25* |
| Power Source Apparatus (PSA-T) | | | | .19* |
| Composite Criterion | | | | .15* |
| Personal History Form | | | | |

Note.— $N = 457$.
* $p < .01$.

TABLE 3

POINT-BISERIAL CORRELATIONS OF PSA-W SCORE WITH THE COMPOSITE CRITERION

| PSA-W raw score | Centile level of dichotomy | r_{pbis} | SE |
|-----------------|----------------------------|------------|-----|
| 19 | 95 | .24 | .09 |
| 12 | 90 | .34 | .07 |
| 11 | 85 | .33 | .07 |
| 10 | 80 | .29 | .06 |

Second, and for two reasons, the product-moment correlation provides a poor estimate of the critical predictor-criterion relationship for these data. This is because (a) the criterion distribution is extremely positively skewed with many values of zero or near zero and (b) it is not the task of the present battery to discriminate throughout the range of scores, but only at its upper end. A priori, it seems clear that whatever fraction of engineers is to be regarded as truly creative, it probably does not exceed 5-20%. Accordingly, a decision was made prior to analysis of the data to examine several biserial cuts at several test score levels versus the continuous criterion. Levels selected were the highest 5%, 10%, 15%, and 20% with the middle two to be regarded as most critical. Scores selected were those on the single best predictor, the PSA-W test. Table 3 contains the results and shows the point-biserial correlation between test and criterion as a function of the level at which the former variable is dichotomized. It will be recalled that a correlation of .33 or .34 is .08 to .09 points above the minimum estimate of Table 2.

Third, the zero-order estimates of the table just cited are really inappropriate because it was originally proposed to construct a battery to be used as a battery, and not as a series of separate subtests. Accordingly, it may be noted that combining the PSA-W and PSA-T with the PHF yields a shrunken R of .28. This represents an increase of .03 points over the best zero-order validity coefficient of .25.

If the effects of the three influences noted were assumed to be additive, they would argue for an estimated correlation of the order of magnitude of .38 to .39 between the combined predictors and the criterion, in Groups

TABLE 4
INTERCORRELATIONS AND VALIDITIES OF
THE LHQ FACTORS

| LHQ factor | 2 | 3 | 4 |
|---------------------------|------|------|--------|
| Socioeconomic background | -.02 | -.07 | -.09 |
| Favorable self-perception | — | -.03 | -.13* |
| Academic achievement | — | — | -.19** |
| Composite criterion | — | — | — |

Note.—*N* = 307. One hundred and fifty cases "held out" to cross validate multiple Rs.
* *p* < .05.
** *p* < .01.

1 and 2 (combined), and with a selection ratio of 10–15%. To evaluate the reasonableness of this estimate a random sample of 100 cases was drawn from the R & D group. The zero-order correlations were corrected for restriction in range; the PSA scores and the PHF score were combined into a single variate through application of the "beta" weights derived from the *R*; and this combined predictor distribution was dichotomized at the level of the top 10% of cases and the top 15% of cases. The resulting validity estimates were .41 and .37, respectively, giving support to the approximate accuracy of the additive estimate.

The second major purpose of this investigation was to identify those personal non-cognitive characteristics (measured by the LHQ) and those environmental characteristics (measured by the JES) which have significantly facilitated or inhibited the expression of creative potential by the present Ss between their initial testing in 1955 and the follow-up of 1964. Data pertinent to this purpose appear in Table 4 which contains the names assigned to the LHQ factors ultimately identified, their intercorrelations, and the correlation of each with the composite criterion. All the factor validities are low, but the identification of Factor 3 and its significant criterion correlation are worth a comment. The responses given by persons scoring high on this factor lead to the following characterization of the respondent: member of an honor society, high ranking in class, a scholarship winner, and a joiner of professional organizations. As indicated by the criterion correlation of -.19, the more creative Ss of this

study could be so characterized a little *less* frequently than their fellows. It should, however, be born in mind that virtually all Ss of Group 1 hold engineering degrees and are professionally employed.

Corresponding data derived from analysis of the JES appear in Table 5. Once more all of the correlations are low, but the criterion validity of Factor 4 sets it apart as a variable of some interest and importance. The typical high scorer on this factor gave item responses implying the following characterization of his job environment: The head of his department publishes, his colleagues hold advanced degrees, the company provides after-hours laboratory facilities for personal research, the head of this department has contributed to patents pending or held, and the head of his department has an M.S. or PhD in engineering. Score on this factor correlates as highly with score on the composite criterion as the best cognitive predictor and suggests the appropriateness of some subsequent discussion of its role and meaning.

The third purpose of this study relates to the task of evaluating the possibility of a Type of Person \times Type of Climate interaction. The procedures described earlier were first employed to establish subsets of Ss with similar profiles on the four JES factors; four subsets were identified. The same methods were then utilized to identify seven subsets of Ss having

TABLE 5
INTERCORRELATIONS AND VALIDITIES OF
THE JES FACTORS

| Factor | 2 | 3 | 4 | 5 |
|--|------|--------|------|-------|
| Utilitarian self-development | -.11 | -.01 | .03 | -.13* |
| Supportive supervisory and peer relationships | — | -.15** | -.14 | .08 |
| Perception of success | — | — | .03 | -.01 |
| Professional and research orientation of supervision | — | — | — | .26** |
| Composite criterion | — | — | — | — |

Note.—*N* = 307. One hundred and fifty cases "held out" to cross-validate Rs.
* *p* < .05.
** *p* < .01.

similar profiles across the three factors of the LHQ. The subsets then became levels of a job climate factor and of a type of person factor, respectively, in an analysis of variance design with two criteria of classification. Table 6 contains the summary of the indicated analysis and reveals that the main effects associated with both sets of subgroups are significant, but that the desired interaction does not even approach significance. To test this result in yet another form, persons in the four quarters of the distribution of scores on the PSA-W were regarded as subsets, constituting four levels of a type of person factor based on creative potential. A second analysis of variance was based on this revised dimension versus type of job climate. This summary appears as Table 7. Once again, no significant interaction emerged, and it was therefore concluded that what constitutes the job environment most conducive to creativity may be generalized across all types of Ss or levels of potential represented in the present R & D group.

DISCUSSION

Before drawing any conclusions on the basis of the data presented, certain limitations in it and in the analytical methods employed must be pointed out.

Restriction in Range

Prominent among the methodological problems of this investigation is that of a conspicuous restriction in the range of talent employed. Clearly, within the available sample, only the R & D engineers of Group 1 had had an equal and substantial opportunity to produce tangible criterion evidence of their creativity. Yet if training and employment

TABLE 6

LHQ VERSUS JES DEFINED SUBGROUPS SUMMARY

| Source | SS | df | MS | F ratio |
|---------------|-----------|-----|--------|---------|
| JES subgroups | 2,309.77 | 3 | 729.92 | 979* |
| LHQ subgroups | 1,870.08 | 6 | 311.68 | 396* |
| Interactions | 1,359.38 | 18 | 75.52 | 96 |
| Error | 30,210.96 | 384 | 78.67 | |

* $p < .01$.

TABLE 7

PSA-W VERSUS JES DEFINED SUBGROUPS SUMMARY

| Source | SS | df | MS | F ratio |
|-----------------|-----------|-----|--------|---------|
| JES subgroups | 1,802.35 | 3 | 600.78 | 7.83* |
| PSA-W subgroups | 1,217.53 | 3 | 405.84 | 5.29* |
| Interactions | 848.33 | 9 | 94.26 | 1.23 |
| Error | 31,996.11 | 417 | 76.73 | |

* $p < .01$.

have partially selected them for aptitude, they are restricted to an unknown extent on the PSA as compared with a population of all engineering graduates. If it is desired to draw even broader inferences regarding the relationship of intelligence to creativity in the general population, the data are obviously inadequate. For example, it may well be that there is some intellectual level beneath which creativity is seldom if ever seen. If so, the two variables are correlated to this extent. Given the present sample of graduates of a difficult and technical college curriculum, however, it is difficult to conceive of any direct test of the hypothesis.

Restriction in another sense exists because only one area of application was considered. No Type of Person \times Type of Environment interaction was discovered. If one type of person is creative in music, another in art, another in literature, and still another in machine design the finding may, clearly, be artifactual. That is, there may indeed be an interaction between type of creative person and type of optimum environment, but no evidence of it in this case because the area of machine design is occupied by only one type of creative person. In short, a limitation of this study is that it does not include the full spectrum of talent; conclusions must be qualified accordingly.

Temporal Interval

A second series of problems intrinsic to these data relates to the time elapsed between testing (1955) and the collection of criterion information (1964). If the interval had been shortened, criterion data would have been less complete and reliable. On the other hand, during a 9-yr period there have, no doubt,

been some true, intrinsic, and differential changes in the creative capabilities of Ss. These, of course, attenuate and lower predictor-criterion relationships in a manner which is for many purposes unwarranted. A case in point may be a postulated differential effect by school attended which would have occurred subsequent to testing during the junior and senior years. Thus, if evidence of short-term validity is desired, the present data probably provide an underestimate of it.

Measurement

Another sort of limitation of the present design is that it involves the mixing of concurrent and predictive data. For example, the JES provides the former and the PSA the latter. To compare their validities implies that the procedures are equally exacting; whereas, it is well recognized that such is not the case.

A second illustration of the same type involves the PHF and LHQ. The former involves only eight open-ended items and has been shown to have predictive validity. When the follow-up was undertaken it was felt that many potential noncognitive predictors of importance had been omitted and the LHQ was constructed accordingly. However, the potential validity of this latter device was surely underestimated, since proven and critical items already included in the PHF were not reintroduced into the LHQ.

Statistics

Since the D^2 statistic was employed as a measure of profile similarity, it seems in order to point out one limitation of this index as used: It involves an unweighted summing of squared deviations across all dimensions of a profile. It was known that these dimensions had differing criterion validities and they might have been differentially effective as potential moderators as well. In any event, a given deviation entered the sum with the same weight regardless of its origin. The implication that each dimension is of the same intrinsic importance in the subgrouping is probably misleading.

Criterion Heterogeneity

Finally, it must be recognized that the Ss of this investigation were employed in a wide

variety of industries, producing many differing products and employing a plethora of methods. It is a truism that the attempt to predict criteria embedded in these widely divergent contexts is fraught with inaccuracy. How much better prediction might be within one company is a matter for conjecture, but the concurrent phase of the present study (1957) suggests that the increment would be very substantial.

INTERPRETATION

An hypothesis suggested by the writer in 1957 is that it is necessary to structure highly or control the association process in the present area of utilitarian creativity in order to enhance or optimize the validity of measurement. For example, the AMT requires only that Ss name machines of any sort within which the given mechanism might function. On the other hand, the PSA test requires that one start with a prescribed power source and produce a prescribed motion sequence.

The entire matter is probably better conceived in terms of current theories of cognitive style (Schroder, Driver, & Streufert, 1967). Adopting this frame of reference, it seems clear that the individual who can accept numerous situational restrictions and still be creative is simply able to integrate more inputs than the one who cannot. Cast in this form the discriminating dimension emerges as one of cognitive complexity.

The present finding of little relationship between mental ability and creativity test results is in accord with the findings of Getzels and Jackson (1962). However, the methodology in the two cases is so divergent as to argue for little real precedent. In the case of academic achievement versus creativity, on the other hand, there is at least an apparently sharp contradiction in outcomes. It is the impression of the writer that this is more apparent than real. Getzels and Jackson (1963) present data which indicate that some cognitive measures beyond the IQ might be useful in the prediction of academic success. However they present no evidence that their creativity tests actually measure any external, normative criterion of creativity of any character. Thus, when they conclude that their

high creativity-lower IQ group does as well in secondary school as their high IQ-lower creativity group, they are, in fact, only comparing selected groups which differ in their performances on two types of cognitive tests. In the present study, on the other hand, score on a cluster of LHQ items implying academic underachievement has been correlated with an external criterion of creativity and found to be significantly associated with it. Here creativity is defined not in terms of test performance, but in terms of such external evidence as patents and patent disclosures. Undoubtedly each type of finding has value, but they are contradictory in name only.

A result of greater importance concerns the relationship of JES Factor 4, professional and research orientation of supervision, to the composite criterion. Going back at least as far as the work of Adamson (1952) on "functional fixedness," there has been considerable, understandable ambivalence regarding the character of leadership most appropriate to an R & D operation and most facilitative of creativity in general. The conflict has had polar opinions ranging from "If we expect people to be creative we can't tell them what to do," on the one hand, to "If we market products within a restricted range, all new ideas do not have equal utility and some guidelines *must* be prescribed," on the other. A resolution has often led to a quite permissive philosophy of leadership. The present data, in another vein, suggest the appropriateness of something quite different. As indicated by the high loading items of Table 5, the optimum environment for the present Ss was one in which they were led by example; the example not only of their head but of their colleagues'. It seems clearly implied that leadership from one who *has done* is not only tolerable, but that it probably constitutes a stimulus to do likewise.

In a related way, it is clear that the treatment of Ss with the optimum environment is far from uniform and that this absence of uniformity tends to confound the relationship of predictors to criterion. Accordingly, adding JES Factor 4 to the multiple correlation of PSA-W and PHF with the composite criterion increases the relationship by .03 correlation

points. This operation is equivalent to asking, "If we knew in advance the sort of environment in which our Ss would function, would a correction for its favorability or unfavorability not enhance prediction?" Both rationally and empirically the answer is "Yes."

Finally, in speaking of the complexity of creativity, it may be of interest to observe the results of combining both predictive and concurrent validity coefficients as though they were comparable and of identifying the three best measures. This task was accomplished by introducing into a "tear-down" multiple regression analysis all 10 potential predictors of the composite criterion of creativity (4 JES factor scores, 3 LHQ factor scores, 2 PSA scores, and score on the PHF). Predictors were then dropped out in inverse order of contribution until only 3 remained. Among these, the largest independent contribution was made by JES Factor 4, professional and research orientation of supervision, the second largest by PSA-T, and the third largest by LHQ Factor 2, favorable self-perception. It is at least intriguing to note that one is a measure of the environment, one a cognitive measure, and one a noncognitive measure. Complex determination of creativity seems clearly implied.

CONCLUSIONS

- 1) The PSA test was found to be a better predictor of the composite criterion of creativity in machine design than a well-regarded mental ability or scholastic aptitude test. The first hypothesis may thus be regarded as supported.

- 2) The PSA test was found to be a better predictor of the present criterion than the less structured AMT. The second hypothesis and the implications of this finding for the relevance of the cognitive complexity dimension, therefore, may be regarded as sustained.

- 3) Since both noncognitive and job environment measures were found to be significant predictors of the present creativity criterion and to enhance prediction when added to that of a cognitive measure, Hypothesis 3 also may be regarded as sustained.

- 4) Since no significant interaction between personal and environmental determiners of

creativity could be detected, Hypothesis 4 must be regarded as rejected in this context.

5) With respect to the predictive validity of the battery of 1955, it is estimated that in a sample of graduates of schools of engineering the best linear combination of measures should correlate .35 to .40 with a composite criterion of their subsequent creativity in machine design. A qualification is that the battery predicts better if scores are realistically dichotomized at a high level than if they are employed as a continuous variate.

REFERENCES

ADAMSON, R. E. Functional fixedness as related to problem solving: A repetition of three experiments. *Journal of Experimental Psychology*, 1952, **44**, 288-292.

BRODIE, W. M. A program to affirm the reality of hierarchical grouping. Unpublished manuscript, Computer Sciences, Purdue University, 1966.

CRONBACH, L. H., & GLESER, G. Assessing similarity between profiles. *Psychological Bulletin*, 1953, **50**, 456-473.

GETZELS, J. W., & JACKSON, P. W. The highly intelligent and the highly creative adolescent: A summary of some research findings. In C. W. Taylor & F. Barron (Eds.), *Scientific Creativity: Its Recognition and Development*. New York: Wiley, 1962.

GUILFORD, J. P. Three faces of intellect. *American Psychologist*, 1959, **14**, 469-479.

KALLICK, M. Organizational determinants of creative productivity. Unpublished doctoral dissertation, Purdue University, 1964.

KLEIN, S. P. Life history and ability correlates of mechanical ingenuity. Unpublished doctoral dissertation, Purdue University, 1965.

MEANS, R. B. Personal and environmental determiners of mechanical ingenuity. Unpublished doctoral dissertation, Purdue University, 1966.

OWENS, W. A., & HENRY, E. R. *Biographical data in industrial psychology: A review and evaluation*. Greensboro, N. C.: The Richardson Foundation, Inc., 1966.

OWENS, W. A., SCHUMACHER, C. F., & CLARK, J. B. The measurement of creativity in machine design. *Journal of Applied Psychology*, 1957, **41**, 297-302.

SCHRODER, H. M., DRIVER, M. J., & STREUFERT, S. *Human information processing*. New York: Holt, Rinehart, & Winston, 1967.

TAYLOR, C. W. (Ed.) *Creativity: Progress and potential*. New York: McGraw-Hill, 1964.

THORNDIKE, R. L. *Personnel Selection*. New York: Wiley, 1949.

TOOPS, H. A. The criterion. *Educational and Psychological Measurement*, 1944, **4**, 271-297.

WARD, J. H., JR., & HOOK, M. E. Application of an hierarchical grouping procedure to a problem of grouping profiles. *Educational and Psychological Measurement*, 1963, **23**, 69-81.

(Received April 8, 1968)

ARCHITECTURE SCHOOL PERFORMANCE PREDICTED FROM ASAT, INTELLECTIVE, AND NONINTELLECTIVE MEASURES¹

CLIFFORD E. LUNNEBORG² AND PATRICIA W. LUNNEBORG

University of Washington

Course grades and faculty ratings through fourth year architecture study were predicted for 228 students from four sets of variables: Architectural School Aptitude Test (ASAT) scores; ASAT scores complemented by 18 traditional academic predictors; the traditional battery alone; ASAT scores complemented by 16 biographic and interest items. ASAT scores alone predicted long-term criteria poorly, but complementing the ASAT with either academic or biographic variables produced the best predictions over all architecture criteria with shrunken validities from .43 to .58. Utility of predictors varied with criteria—faculty ratings were largely determined by traditional intellectual measures while design performance was a function of nonintellectual and background information which appears essential to prediction in areas of divergent thinking.

This study is part of a continuing search for measures of divergent thinking and for better predictors of performance in occupational areas depending on such ways of thinking. The traditional predictors of college performance, that is, high school GPA and tests of verbal and quantitative aptitude, have always worked much better estimating success in English, mathematics, and biology courses than they have in art, music, and architecture. For this reason, the construction of the Architectural School Aptitude Test (ASAT; Educational Testing Service, 1965) centered around the predictive effectiveness of traditional measures versus tests designed specifically to tap abilities which architects had judged were related to success in architecture school.

The original validity study with the ASAT indicated that it did not outperform high school rank in class or GPA, but was a useful addition to high school record in predicting architecture performance (Pitcher, Olsen, & Solomon, 1962). Further, evidence was presented that traditional verbal and mathematics scores in combination with high school record were inferior to the ASAT-high school record combination in predicting first year archi-

itecture GPA. Even ignoring high school record the verbal and mathematics tests were not as predictive as the six ASAT subtests (adjusting for shrinkage).

The present study was prompted by two effects of the high rate of attrition among students in the validation study (only 24% or 145 students had completed their studies in 5 yr.). First, the small size of samples at the 12 participating schools made the results somewhat inconclusive. Secondly, predictors were consequently judged primarily in terms of first year architecture GPA; long-term criteria such as completion or non-completion for academic reasons were necessarily slighted. It was felt that additional evidence of validity for the ASAT, traditional and nontraditional (nonintellectual) measures, was needed over a range of criteria of architecture school success.

METHOD

Subjects. The total sample consisted of 228 students entering the University of Washington School of Architecture between 1964 and 1966. This group was predominantly male (92%), single (96%), and from Washington state high schools (85%).

Predictors and criteria. The initial pool of predictors consisted of age and sex, and (a) ASAT total and six part scores (interest vocabulary, sensitivity to visual phenomena, science reasoning, intersections, complex space fitting, and incorporated lines), (b) six cumulative high school GPAs: English, foreign languages, mathematics, natural sciences, social studies, and full-credit electives, (c) 10 tests: ACE Psychological Exam (Quantitative), Guilford-Zim-

¹ A modified version of this paper was presented at the American Psychological Association meeting, San Francisco, September 1968.

² Requests for reprints should be sent to Clifford E. Lunneborg, Director, Bureau of Testing, University of Washington, Seattle, Washington 98105.

TABLE 1

CORRECTED MULTIPLE CORRELATION COEFFICIENTS FOR BEST SETS OF PREDICTORS OF
SEVEN CRITERIA OF ARCHITECTURE SCHOOL SUCCESS

| Predictor set | Criteria | | | | | | |
|------------------------------------|---------------------------|----------------------------|---------------------------|----------------------------|-----------------------|----------------------------|------------------------------|
| | First year arch GPA | Second year arch GPA | Third year arch GPA | Fourth year arch GPA | Arch design GPA | All-uni- versity GPA | Average faculty rating |
| ASAT total and six part scores | 35(4) | 38(3) | 14(1) | −06(1) | 20(1) | 29(2) | 43(2) |
| 18 traditional predictors | 37(5) | 33(3) | 56(6) | 40(5) | 33(3) | 48(4) | 52(6) |
| ASAT, 18 traditional predictors | 46(10) | 43(4) | 58(6) | 48(7) | 43(3) | 48(4) | 55(8) |
| ASAT, 16 nonintellective variables | 48(8) | 48(7) | 44(5) | 46(5) | 52(9) | 38(7) | 54(7) |
| <i>N</i> | 226 | 201 | 147 | 78 | 124 | 228 | 166 |

Note.—Number of variables in best set follows R_c in parentheses. Decimal points have been omitted.

merman Survey—Part I (verbal comprehension), CEEB intermediate mathematics, Washington Pre-College (WPC) tests of English usage, spelling, reading speed, reading comprehension, mechanical reasoning, spatial ability, and applied mathematics, and (d) 50 biographic and interest variables derived from admission applications or from a questionnaire administered in introductory architectural design.

There were seven criteria: first year architecture GPA (5 quarter hr. of introductory architecture and 9 hr. of drawing), second year architecture GPA (18 hr. of architectural design and 6 hr. of water color), third year architecture GPA (18 hr. of architectural design and 24 hr. of technical architecture), fourth year architecture GPA (18 hr. of design and 29 hr. of technical architecture), architecture design GPA (design beyond second year), cumulative all-university GPA, and the average rating (5-point scale) by three architecture professors of student potential based on personal interviews in the second year of architecture.

Procedure. Intercorrelations among the 75 predictors and seven criteria were the basis for narrowing down the number of variables for four sequential predictor selection analyses: ASAT total and part scores; age, sex, high school GPAs and test scores; ASAT with these 18 traditional predictors; ASAT with 16 of the original 50 nonintellective measures. In each of the sequential predictor selections (Horst & Smith, 1950) variables were added to the predictor set as long as their contribution to prediction outweighed the expected shrinkage in multiple correlation owing to increased number of predictors. No limit was placed on the potential number of predictors to be chosen so that as many useful variables would be identified as possible. Because of the fluctuation inherent in multiple correlations from one group to another, especially if groups are small, multiple correlation coefficients reported here have been corrected (R_c), that is, reduced to reflect the expected between-sample shrinkage owing to sample size and number of predictors.

RESULTS AND DISCUSSION

The mean ASAT total score for the entire group was 567; $SD = 101$. The average student entered architecture approximately 1 yr. after graduating from high school and 24% were enrolled in some other college prior to entering the university.

The simple correlation coefficients for ASAT total score with all criteria compared closely to the multiple R_c 's when ASAT total and part scores were reweighted to provide the best prediction. It thus appears that the original weighting devised for the ASAT is broadly applicable. The criteria based on third and fourth year work as well as the design GPA, however, had validities of only .18 and below with ASAT total score. In the prediction of faculty ratings slightly better predictions were obtained by increasing the weight given two of the six parts, interest vocabulary and science reasoning. R_c values for the reweighted ASAT parts are given in Table 1.

As can be seen from Table 1, for all criteria except first and second year architecture work, the traditional battery (age, sex, high school GPAs, and 10 tests) provided substantially better predictions than ASAT scores. However, ASAT scores complemented either with the traditional battery or with the biographic and interest variables performed better than the traditional battery alone for all criteria save all-university GPA where prediction from social studies and natural science GPAs and English usage and

TABLE 2

STANDARD PARTIAL REGRESSION WEIGHTS FOR BEST SETS OF PREDICTORS OF
SEVEN CRITERIA OF ARCHITECTURE SCHOOL SUCCESS

| Predictors | First year arch GPA | Second year arch GPA | Third year arch GPA | Fourth year arch GPA | Arch design GPA | All-uni- versity GPA | Average faculty rating |
|--|---------------------------|----------------------------|---------------------------|----------------------------|-----------------------|----------------------------|------------------------------|
| ASAT predictors | | | | | | | |
| ASAT total score | 32(1) | 42(1) | | | | | |
| ASAT Part I (interest vocab) | 14(2) | | | | | | 36(1) |
| ASAT Part II (sensitivity to phen) | | 09(3) | | | | 13(2) | |
| ASAT Part III (science reasoning) | | -13(2) | | | | 24(1) | 15(2) |
| ASAT Part IV (intersections) | | | 17(1) | -13(1) | 24(1) | | |
| ASAT Part VI (complex space fitting) | -07(3) | | | | | | |
| ASAT Part VII (incorporated lines) | -06(4) | | | | | | |
| R_c | 35 | 38 | 14 | -06 | 20 | 29 | 43 |
| 18 traditional predictors | | | | | | | |
| Sex (male) | -11(5) | | | | | | |
| Age | | | | | | | 16(3) |
| HS English GPA | | | | 44(1) | | | |
| HS mathematics GPA | | | | 36(3) | | | |
| HS social studies GPA | 20(2) | | 47(1) | -46(4) | 37(1) | 26(2) | |
| HS natural science GPA | | 3(1) | | | | 17(1) | 30(2) |
| HS electives GPA | | | -13(5) | | | | -10(4) |
| Guilford-Zimmerman verbal comp | | | -33(3) | | -20(2) | | |
| CEEB intermediate mathematics | | | 25(2) | -32(2) | | | |
| WPC English usage test | | | | | | 13(4) | 28(1) |
| WPC spelling test | 12(3) | | 20(4) | | | | |
| WPC reading comprehension test | | 16(2) | | | | | 15(6) |
| WPC mechanical reasoning test | 36(1) | | | -17(5) | | 16(3) | |
| WPC spatial ability test | -14(4) | | 11(6) | | | | |
| WPC applied mathematics test | | 14(3) | | | 13(3) | | -10(5) |
| R_c | 37 | 33 | 56 | 40 | 33 | 48 | 52 |
| ASAT plus 18 traditional pre- dictors | | | | | | | |
| ASAT total score | 34(1) | 43(1) | | | | | |
| ASAT part I (interest vocab) | 21(6) | | | | | | 24(1) |
| ASAT part II (sensitivity to phen) | | | | | | | -13(7) |
| ASAT part III (science reasoning) | | -20(3) | 19(4) | 40(6) | | | |
| ASAT part IV (intersections) | | | | | 30(2) | 16(3) | 16(4) |
| ASAT part VII (incorporated lines) | -09(10) | | | | | | |
| Sex (male) | -14(7) | | | | | | -11(6) |
| Age | 15(2) | | | | | | |
| HS English GPA | 15(3) | | | 39(1) | | | |
| HS mathematics GPA | | | | 42(3) | | | |
| HS social studies GPA | | | 50(1) | -44(4) | 43(1) | 23(2) | |
| HS natural science GPA | | 21(2) | | | | 22(1) | 29(2) |
| HS electives GPA | | | -12(6) | | | | -11(5) |
| Guilford-Zimmerman verbal comp | -33(5) | | -35(3) | -23(7) | -26(3) | | |
| CEEB intermediate mathematics | | | 24(2) | -41(2) | | | |
| WPC English usage test | 12(9) | | | | | 11(4) | 17(3) |
| WPC spelling test | | | 19(5) | | | | |
| WPC reading comprehension test | | 08(4) | | | | | 09(8) |
| WPC mechanical reasoning test | 15(8) | | | -35(5) | | | |
| WPC spatial ability test | -18(4) | | | | | | |
| R_c | 46 | 43 | 58 | 48 | 43 | 48 | 55 |
| ASAT plus nonintellective variables | | | | | | | |
| ASAT total score | 34(1) | 40(1) | | | | | |
| ASAT Part I (interest vocab) | | | | | | | 34(1) |

Table 2—(Continued)

| Predictors | First year arch GPA | Second year arch GPA | Third year arch GPA | Fourth year arch GPA | Arch design GPA | All-uni- versity GPA | Average faculty rating |
|--------------------------------------|---------------------------|----------------------------|---------------------------|----------------------------|-----------------------|----------------------------|------------------------------|
| ASAT Part II (sensitivity to phen) | | 13(6) | | | | 14(5) | |
| ASAT Part III (science reasoning) | | −16(5) | | | | 22(1) | 14(6) |
| ASAT Part IV (intersections) | | | | | 27(2) | | |
| ASAT Part VII (incorporated lines) | −09(6) | | 17(5) | | | | |
| Father's occupational level (Roe) | −22(2) | −26(2) | | | | −14(4) | |
| Father college graduate | | | | 45(1) | | | |
| Mother employed outside home | | | | | −17(3) | | 12(5) |
| Mother college graduate | | | | −32(3) | −21(5) | | |
| Firstborn (including onlies) | −11(5) | | | | | | |
| Interval HS to entrance in arch | 13(8) | | | | | | |
| Attended HS in state | −09(4) | −11(4) | | 14(5) | −15(9) | −17(2) | −19(2) |
| HS honor recipient | | | 27(1) | | 18(4) | | 16(3) |
| Part-time job in college | | | 29(3) | | 16(8) | 18(3) | −10(7) |
| Architecture HS vocational choice | 14(7) | | 26(2) | | 31(1) | | |
| Father's occupation business contact | | | −20(4) | −28(2) | −21(6) | −12(6) | |
| Father's occupation technical | | 18(3) | | | −20(7) | | |
| Creative people cited in art, arch | 17(3) | | | 15(4) | | | |
| Service motivation for architecture | | 11(7) | | | | 09(7) | 15(4) |
| R _c | 48 | 48 | 44 | 46 | 52 | 38 | 54 |

Note.—Order of selection in parentheses following weights with decimal points omitted. Predictor intercorrelations based on 228 Ss administered ASAT of whom 166 had Washington Pre-College (WPC) scores and high school (HS) grades, and 186 had biographic data. Table includes only predictors selected at least once.

mechanical knowledge tests could not be improved upon. The ASAT together with the traditional battery provided the best predictions of faculty rating, third, and fourth year grades, while biographic and interest items combined with the ASAT provided the highest multiple correlations with first and second year grades as well as advanced design. Table 2 reports order of selection and standard partial regression weights for the variables in each predictor selection.

Briefly, the biographic correlates of architecture performance based on the predictor selections involving ASAT and nonintellective variables include the following. Roe's (1956) occupational level of father was often selected and indicates that the higher family socio-economic status, the better will be student performance in architecture. Similarly, for fourth year grades, father's education was the most potent predictor of all. A very good addition to prediction was having attended secondary school out of the state, and perhaps this variable too reflects socio-economic status through capacity to pay nonresident tuition and campus living costs. Performance in architecture was aided by having received

honors in high school, by deciding in high school on a vocation in architecture, and curiously, by holding a part-time job in college. Choosing architecture from a social service motivation especially contributed to faculty opinion of student potential. The last nonintellective variable of consequence was that of father's occupation in business contact and selling (Roe, 1956) which adversely affected several criteria.

A first conclusion from examining Table 1 is that given the uneven predictability of criteria within a single school of architecture, probably any school wishing to use the ASAT must conduct its own validation study, selecting and weighting variables which reflect the emphases in its particular curriculum. The relative importance to success of design courses, technical courses, and courses required in areas outside architecture, such as physics and social science, will determine the kinds of predictors that get selected.

To illustrate this point from the present study, faculty ratings of student potential were best estimated from the interest vocabulary and science reasoning parts of the ASAT, high school natural science GPA, and

WPC English usage. Remembering that verbal and mathematics tests were excluded from the final ASAT battery on the grounds that they overlapped with interest vocabulary and science reasoning (Pitcher et al., 1962), faculty ratings would appear solely a function of traditional, intellectual predictors. Advanced design course performance, on the other hand, emphasized in its prediction one of the performance subtests of the ASAT, intersections, and a number of biographic and interest variables: early interest in architecture, receipt of honors in high school, father employed in something other than selling or a technical occupation, mother not employed outside the home. All-university grade average, depending in part on nonarchitecture course work required for graduation, was best predicted by the traditional "classic" battery of measures of academic aptitude and achievement.

Although choice of criterion influenced the effectiveness of all predictors including the ASAT and its parts, this study provides additional evidence of the usefulness of the ASAT as a tool for guiding or advising prospective architecture students. It appears, however, that the effectiveness of the ASAT

would be considerably reduced were it not supplemented with other intellectual measures or with biographic data. A cautious generalization is that where criteria are short-term, augmentation with traditional predictors works well, but where criteria approach the ultimate in terms of architecture success, nonintellectual background and interest variables account for significant variance in addition to the ASAT. For some time all architectural criteria should be considered equally important. At this stage of exploring divergent thinking and its occupational counterparts, it is as critical to know how an individual will fare in his first year of study as it is to know whether he succeeds professionally some years hence.

REFERENCES

- EDUCATIONAL TESTING SERVICE. *Architectural School Aptitude Test: A guide to interpretation of scores*. Princeton: Educational Testing Service, 1965.
- HORST, P., & SMITH, S. The discrimination of two racial samples. *Psychometrika*, 1950, 15, 271-289.
- PITCHER, B., OLSEN, M., AND SOLOMON, R. A study of the prediction of academic success in architectural school. Princeton: Educational Testing Service, 1962.
- ROE, A. *The psychology of occupations*. New York: Wiley, 1956.

(Received April 15, 1968)

RELATIONSHIP BETWEEN INTERVIEW STRUCTURE AND INTERINTERVIEWER RELIABILITY IN AN EMPLOYMENT SITUATION

DONALD P. SCHWAB¹ AND HERBERT G. HENEMAN III²

Graduate School of Business, University of Wisconsin

Eighteen experienced job interviewers were assigned randomly to three groups which differed in the degree of interview structure. The structured group was not permitted to deviate from a predetermined interview format. The same format was used by the semistructured group and deviations were permitted. The unstructured group was free to interview applicants in any manner. A job specification for the job of clerk-stenographer was developed. On the basis of this specification, descriptions of five hypothetical job applicants were constructed. A separate description was given to each of five female undergraduates who were instructed in their job applicant roles. Each group interviewed and then ranked the five job applicants. The amount of interinterviewer agreement within groups was found to be positively related to the degree of interview structure.

Three major reviews of the selection interview literature have concluded that structured interviews are probably superior to less structured ones (Mayfield, 1964; Ulrich and Trumbo, 1965; Wagner, 1949). Mayfield (1964) found that published interview studies reporting the greatest interrater agreement tended to employ structured interview formats. Ulrich and Trumbo (1965) and Wagner (1949) concluded that the greatest interview reliability and validity resulted when structured interviews were used. These general findings, however, are not without exception. For example, Hakel (1966) employed a highly structured interview design in an employment situation, yet concluded that interrater "reliability correlations . . . showed no more agreement than did Scott's interviewers in 1915 [p. 45]."

Furthermore, any of these general conclusions are of dubious value because they have been based on comparisons between studies which leave uncontrolled a number of variables known to influence interinterviewer reliability. Specifically, studies have shown that interviewers' decisions are related to the

homogeneity between applicants, that is, the variance of total qualifications in the applicant groups presented for appraisal (Carlson and Mayfield, 1967; Sydiaha, 1958), order of applicant presentation (Rowe, 1967), and the type of rating form employed to record interviewer judgments (Carlson and Mayfield, 1967). Other variables suspected of influencing interinterviewer reliability include length of interview and type of job (Mayfield, 1964), and interviewers' experience and knowledge of the job (McMurry, 1947). These and probably other uncontrolled variables introduce an indeterminate amount of error into cross-study comparisons of interview reliability.

Despite the obvious need to investigate the impact of interview structure under more controlled conditions, Mayfield (1964) stated "in no study located was the amount of structure varied systematically to see what effect this would have on the results [p. 242]." The present study represents a preliminary attempt to close this gap by investigating the effect of three degrees of interview structure on interinterviewer reliability. Through the research design an effort was made to control for differences in rating forms, informational cues other than those called for in the degree of structuralization, order of applicant presentation, homogeneity of applicants' characteristics, interviewers' knowledge of the

¹ Requests for reprints should be sent to Donald P. Schwab, Graduate School of Business, University of Wisconsin, 1155 Observatory Drive, Madison, Wisconsin 53706.

² The authors gratefully acknowledge the assistance of Charles S. Bridgman throughout the course of the research.

job, and interviewing experience. The research hypothesis stated that interinterviewer reliability is positively related to the degree of interview structure.

METHOD

Subjects

Eighteen male assistant city department managers participating in a management development program served as Ss for the study. The Ss, who interviewed employment applicants as a normal part of their job duties, were assigned randomly to one of three interview groups: structured, semistructured, and unstructured ($n = 6$ per group). Comparison of the three groups on length of service in present job ($M = 6.6$ yr.) and number of employment interviews conducted ($M = 22.3$ per yr.) showed no significant differences.

Five volunteer female undergraduates served as applicants for the job of clerk-stenographer employed by the city government. Each of the girls had previous experience as an interviewee.

Procedure

The job of clerk-stenographer was chosen for the study because Ss were generally familiar with that position. A detailed clerk-stenographer job specification was developed describing the nature of the work, essential knowledge and abilities, and necessary training and work experience. This specification closely paralleled the actual clerk-stenographer job specification used by the city. A job application form also was prepared which provided for information on biographical data, education, work experience, tested abilities for the job, outside interests and activities, and stated interest in the job.

A hypothetical description, which included a completed application blank for each girl, was developed for each of the five participating applicants. All applicants were described as meeting the minimum qualifications for the position required by the job specification, though their qualifications varied above the minimum. In addition, the hypothetical descriptions varied in terms of age, marital status, place of residence, outside interests and activities, and interest in the job.

An effort was made to develop low homogeneity among applicants in order to insure a fairly high degree of agreement among interviewers in the structured group. A pretest was conducted by giving judges not associated with the final study the five written applications, with instructions to rank them from best to poorest. Because these original descriptions elicited low agreement between judges, they were rewritten to increase the heterogeneity of the information contained on the application blanks. In most cases, this involved an alteration of the applicants' tested abilities for the job, though these never were lowered below the necessary hiring minimum.

After revision, a hypothetical description was given to each applicant for memorization. In addition to the information on the application blank, each girl was given other information which dealt with home and family life, future plans, and more detailed work experience data to complete their hypothetical descriptions. The applicants were informed about the interviewing procedures they would experience and were told of the importance of providing consistent answers to similar questions between interviewing groups.

All Ss were given copies of the detailed clerk-stenographer job specification and told that they would interview five applicants for that position. Although Ss did not know the purpose of the study, they were told that the applicants were in reality college students and that the information they obtained from the applicants was written specifically for this study. In addition, Ss were told to assume that the applicants had passed the necessary tests and interviews in the centralized personnel office and had been referred to Ss for a hiring interview. This procedure was identical to the selection procedure used in the city government. Finally, Ss were told that each interview group had a maximum of 10 min. to interview each applicant, that after completing all of the interviews they should rank the five girls as potential clerk-stenographers from best (1) to poorest (5), and that they should not discuss the applicants between interviews or during the ranking process.

Each group then was assigned a separate room and given further specific instructions. (a) *Structured group*: The Ss were given blank copies of the prepared application form and instructed to ask each applicant factual questions pertaining only to the information called for on the application form. (b) *Semistructured group*: The Ss were also given blank copies of the same form and were told to obtain all information called for on the application form. In addition, this group was told that they could ask each applicant any additional or follow-up questions which appeared appropriate. (c) *Unstructured group*: The Ss were not given application forms. Rather, they were told that they could ask the applicants any questions during the time allotted for each interview.

A time schedule was drawn up for the interviews, and the applicants were assigned randomly to the three interview groups and the five time periods. Kendall's coefficient of concordance was used to determine the amount of agreement between Ss' rankings within each of the three groups (Siegel, 1956).

RESULTS

The results are presented in Tables 1 and 2. Table 1 shows interviewer rankings of the applicants within each group. It is apparent that the greatest degree of interrater agreement exists in the structured group. Only Applicant C obtained as consistent rankings

TABLE 1

APPLICANT RANKINGS BY INTERVIEWERS WITHIN EACH GROUP

| Job applicants | Interviewers | | | | | | | | | | | | | | | | | |
|----------------|--------------------------|---|---|---|---|---|------------------------------|---|---|----|----|----|----------------------------|----|----|----|----|----|
| | Structured group (n = 6) | | | | | | Semistructured group (n = 6) | | | | | | Unstructured group (n = 6) | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| A | 3 | 5 | 5 | 5 | 3 | 5 | 1 | 3 | 2 | 5 | 3 | 3 | 2 | 5 | 2 | 3 | 5 | 4 |
| B | 2 | 3 | 2 | 2 | 2 | 2 | 3 | 1 | 4 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 3 |
| C | 5 | 2 | 3 | 3 | 4 | 3 | 4 | 5 | 3 | 3 | 4 | 4 | 3 | 1 | 4 | 1 | 3 | 1 |
| D | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 4 | 1 | 2 | 1 | 1 | 5 | 3 | 3 | 4 | 4 | 2 |
| E | 4 | 4 | 4 | 4 | 5 | 4 | 5 | 2 | 5 | 4 | 5 | 5 | 4 | 4 | 5 | 5 | 1 | 5 |

in the semistructured or unstructured groups. A statistical summary of the data is presented in Table 2. As hypothesized, the amount of agreement between interviewers increased as the degree of interview structure increased. While to our knowledge it is not possible to test the significance of the difference between groups for the coefficient of concordance, the differences in the magnitude of the coefficients are in the hypothesized direction.

DISCUSSION

The findings obtained in the present pilot investigation are important because they represent the first known effort to investigate systematically the impact of the degree of interview structure on interinterviewer reliability. There are, however, several factors in addition to the small sample size which should be considered before generalizing from the results obtained here.

First, attention must be directed to the fact that applicant qualifications were deliberately manipulated to ensure a moderate degree of interinterviewer agreement in the structured

group. If there is interaction between applicant homogeneity and interview structure, the results obtained could be varied solely by changing the homogeneity between applicants. Specifically, if little difference exists between applicants in terms of job qualifications, low interrater agreement might well be obtained in all groups. Alternatively, applicants might be so heterogeneous that interinterviewer agreement in the structured interview might be close to 1.0. There thus would be little opportunity for higher reliabilities to be obtained in the less structured interview groups.

An examination of the interinterviewer agreement obtained (Table 2) suggests that neither extreme occurred. While interviewer agreement in the structured group was fairly high, it was not perfect. Thus, the possibility existed for greater interinterviewer agreement in the less structured groups. The authors deliberately sought to enhance this possibility by withholding some questions from the structured and semistructured format which elicited relevant information about the applicants. In fact, greatest interinterviewer agreement occurred in the structured group. It thus appears that when there is little homogeneity between applicants, the degree of interview structure has an impact on the reliability of interview rankings.

Major needs for future research on the determinants of interinterviewer reliability are suggested by this finding. Subsequent research on interview structure should vary the degree of homogeneity between applicants. Conversely, investigators concerned with the impact of varying degrees of applicant homo-

TABLE 2

KENDALL'S COEFFICIENT OF CONCORDANCE
FOR EACH INTERVIEW GROUP

| Item | Structured group | Semi-structured group | Un-structured group |
|----------------------------|------------------|-----------------------|---------------------|
| Coefficient of concordance | .79** | .43* | .36 |

* $p < .05$.
** $p < .01$.

geneity should investigate its possible interactive effects with interview structure.

Finally, it should be remembered that the interviewers ranked the applicants and that ties were not permitted. There is evidence suggesting that ratings tend to result in a higher degree of interinterviewer reliability than rankings because each applicant is rated independently and because ties are possible (Carlson and Mayfield, 1967). The amount of agreement within each group thus may be understated, relative to the amount of agreement possible with the use of ratings. The actual comparison of the agreement between rankings and ratings over varying degrees of interview structure remains to be tested.

In summary, on the basis of the sample and technique employed, it appears that the degree of interview structure may have a significant impact upon the degree of interinterviewer agreement. The limited number of judges and applicants suggests the need for replication of this finding. Furthermore, additional research is required to determine possible interaction between degrees of applicant homogeneity and interview structure.

REFERENCES

- CARLSON, R. E., & MAYFIELD, E. C. Evaluating interview and employment application data. *Personnel Psychology*, 1967, **20**, 441-460.
- HAKEL, M. D. Perceiver differences in interpersonal perceptions. An analysis of inter-rater agreement on scaled-expectation rating scales in an employment interview setting. Unpublished doctoral dissertation, University of Minnesota, 1966.
- MAYFIELD, E. C. The selection interview: A re-evaluation of published research. *Personnel Psychology*, 1964, **17**, 239-260.
- McMURRY, R. N. Validating the patterned interview. *Personnel*, 1947, **23**, 2-11.
- ROWE, P. M. Order effects in assessment decisions. *Journal of Applied Psychology*, 1967, **51**, 170-173.
- SIEGEL, S. *Nonparametric Statistics*. New York: McGraw-Hill, 1956.
- SYDIAHA, D. The relation between actuarial and descriptive methods in personnel appraisal. Unpublished doctoral dissertation, McGill University, 1958.
- ULRICH, L., & TRUMBO, D. The selection interview since 1949. *Psychological Bulletin*, 1965, **63**, 100-116.
- WAGNER, R. F. The employment interview: A critical summary. *Personnel Psychology*, 1949, **2**, 17-46.

(Received April 18, 1968)

AUTOMATED INSTRUCTION FOR VIGILANCE TRAINING¹

DENNIS A. ATTWOOD

AND

EARL L. WIENER²

*Defence Research Establishment Toronto
Downsview, Ontario, Canada*

*Department of Industrial Engineering
University of Miami*

The performance of Ss trained in a visual monitoring task with an auto-instructional device was compared with that of Ss trained by practice alone. The experimental group had three 50-min. training sessions on a device which included the standard monitoring task, but allowed *S* to select his signal schedule and call for immediate knowledge of results, or signal cueing (prompting), or both, and to test himself with no training aids available. Subsequent testing on the standard task revealed that Ss trained with auto-instruction showed a much higher detection rate ($p < .01$) than the control group, with no increase in commissive errors. Reasons for the success of auto-instruction in vigilance training are discussed.

Vigilance, or monitoring displays for difficult-to-detect, low-probability signals, has been studied extensively since World War II, but only recently have investigators turned to training techniques. Mere practice at a vigilance task results in no more than slight improvement in detections (Wiener, 1968), but several *Es* have been successful in obtaining positive transfer effects when Ss were first trained with immediate knowledge of results (KR) and then transferred to a condition where they were deprived of this feedback information (Adams & Humes, 1963; Hardesty, Trumbo, & Bevan, 1963; Mackworth, 1964; Wiener, 1963, 1967). Recently another training aid, known variously as prompting, cueing, or alerting, whereby the operator is warned that a signal is soon to appear, has been employed, but with mixed success (Annett & Paterson, 1967; Colquhoun, 1966). The use of KR and cueing as training aids for perceptual tasks has been reviewed by Aiken and Lau (1967). Wiener and Attwood (1968) used combinations of the two in a factorial experiment and found KR highly beneficial to transfer of training, but cueing ineffective.

Recent research in autoinstructional techniques outside of the monitoring field, employ-

ing devices ranging from simple teaching machines to elaborate computer-based systems, suggests some possibilities for training monitors. While the results of the transfer of training experiments previously cited have been promising from a practical point of view, highly regimented training programs may be less effective than autoinstructional techniques in which the trainee would have some options. By implementing a flexible program with options on the training aids, *S*'s performance may be enhanced for one or more of the following reasons: (a) *S* may actually know, or think that he knows, the best way to train himself; (b) *S*'s motivation may be heightened when he is an "active partner" in his own training, rather than a passive recipient of a rigid training regime; (c) autoinstruction allows elements of self-pacing and self-testing.

There is currently no experimental evidence on the use of self-training devices for monitoring tasks. The closest to it are the experiments of Swets, Millman, Fletcher, and Green (1962), Swets, Harris, McElroy, and Rudloe (1966), and Weisz and McElroy (1964), who used computer-aided techniques, including KR, prompting, and others, in the identification of multidimensional aural and visual patterns. They found that Ss trained in accordance with the usually accepted principles of autoinstruction did not show significantly better results than those given simple stimulus presentations. Also, in the second Swets et al. study (1966), Ss had a variety of options available, but their performance was no

¹ This investigation was supported in whole by Public Health Service Research Grant No. UI00014 from the National Center for Urban and Industrial Health. It is based on a thesis by the first author, submitted in partial fulfillment of requirements for the M.S.I.E. degree at the University of Miami.

² Requests for reprints should be sent to Earl L. Wiener, 202 Engineering, University of Miami, Coral Gables, Florida 33124.

better than that of Ss in the previous experiment (1962), where they had received a fixed regime.

In this study an experimental group performed, on five successive days, a pretest, three sessions of training on the autoinstructional device, and a posttest. A control group performed all five sessions under "standard conditions" of visual monitoring, which were identical to the pretest and posttest of the experimental group. The autoinstructional device allowed Ss in the experimental group options over their use of KR, cueing, the signal rate and distribution, and provided the option for self-test periods without training aids.

METHOD

Subjects

The Ss were 34 male undergraduates from the University of Miami, with no previous experience in the monitoring task. They were recruited by means of posters displayed on campus and were paid \$10 upon completion of the 5-day experiment.

General Apparatus Features

The monitoring task for both groups was the detection of an abnormally large deflection of a voltmeter needle. The meter faceplate was painted flat black and the needle was white. The nonsignal stimulus was a 20° rightward needle deflection, while the signal was a 30° deflection. Both the signal and nonsignal stimuli were the result of electrical discharges applied to the voltmeter through a resistor-capacitor circuit, triggered from a coded punch paper tape. In both the standard vigilance program and the training program, tape readers produced 50 stimuli/min.

Standard Apparatus

Room 1 was partitioned into three booths, in which Ss viewed the voltmeter via closed-circuit television. To achieve auditory isolation Ss wore earphones which played white noise. The Ss indicated the presence of a signal by pressing a silent pushbutton switch, mounted on the table in front of them. Responses were recorded on an Esterline-Angus event recorder, and also on two banks of reset event counters that recorded the number of missed signals and sorted S's responses into detections and false alarms.

Training Group Apparatus

The apparatus in Room 2 consisted of a display board, a selection panel, a silent pushbutton response switch, and earphones. The display board was a vertical black panel mounted 30 in. in front of S. It contained the recessed meter display, a set of

TABLE 1
INTER SIGNAL INTERVALS (IN SECONDS) FOR
SIGNAL RATE OPTIONS

| Signal distribution | Signal rate | | |
|---------------------|-----------------|-----------------|------|
| | Slow | Medium | Fast |
| Fixed | 36 | 24 | 12 |
| Random | 72 ^a | 30 ^a | — |

^a Average intersignal interval.

three color-coded KR lights, information lights for the rest and selection periods, and four one-plane digital readouts which provided summary feedback information to S. The selection panel contained 12 momentary-contact switches mounted on a plastic, translucent faceplate. These illuminating switches were arranged into five vertical columns which corresponded to the five choices in the TRAIN mode. The monitoring task per se was the same as the standard task. The selection panel operated a relay logic bank which gated the correct signal and cueing information from punched paper tape to the display. Performance data were collected as in the standard task. The option choices were recorded on 15 pens of the Esterline-Angus recorder.

The earphones were connected through an intercom system to three separate audio sources. A 400 cps recorded tone was used as a background noise when the branching mode was activated, and a 1000 cps tone provided the alerting signal during the cueing option. At all other times, except during the rest period, white noise was used for auditory isolation; during the rest period the earphones were silent.

Signal Schedules

The standard monitoring task provided 32 signals per session, spaced in a different random order in each of the five daily 48-min. sessions. The 32 signals were equally divided into four 12-min. periods within each session, with the minimum intersignal interval being 18 sec. In the training sessions, three signal rates and two signal distributions were available as options to S. The signal frequencies are listed in Table 1. The minimum intersignal interval was 9.6 sec. The intersignal intervals for the random distributions were determined from a table of uniformly distributed random numbers. The RANDOM SLOW signal rate provided either 4, 5, or 6 signals per 6-min. period.

Procedure

The Ss were randomly assigned to two equal groups, Control (C) and Training (T). On the first day all Ss monitored for a 48-min. session on the standard apparatus. The Ss were read standard instructions, including a demonstration of five signals spaced 10 nonsignals apart.

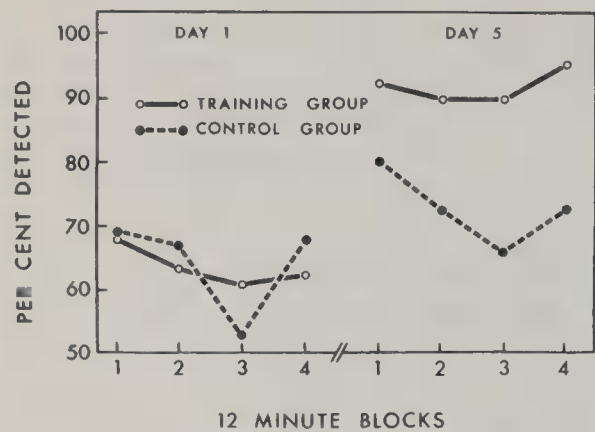


FIG. 1. Percentage of signals detected as a function of length of watch by training and control groups on pretest and posttest days.

Since Group T monitored on both the standard and training equipment, they were told that the change of apparatus was to determine how well they performed on Day 5 after receiving three days of training. To cancel any bias due to these instructions, Ss in Group C were also told that the results of the Friday session would determine the effects of practice, and therefore these posttest results were what counted. For the next three days Group C returned to the same task. On the last day, Ss in both groups monitored the standard task for 48 min.

Autoinstruction

On the second day each S in Group T monitored in Room 2. Recorded instructions explained the use of the equipment. For each 6-min. period, S could choose either the TRAIN or TEST mode. He was allowed a 1-min. rest between periods and 20 sec. to choose the next program. The TRAIN mode allowed S free choice at all times of five signal rates, KR, and cueing, as well as a choice of whether to receive the corrective branching option. Choice of the KR option presented S with a green detection light as long as the response switch was depressed within 2.5 sec. following a signal. When the response switch was pressed outside of this 2.5 sec. interval, a red false alarm light was illuminated. If no response was made within the 2.5-sec. interval following a signal, the amber missed-signal light was illuminated for 2 sec. By choosing the CUE option, S was presented with a 1000 cps. tone 2.4 sec. in duration beginning 3.4 sec. before the onset of each signal. These aids were essentially the same as those employed in a previous study (Wiener & Attwood, 1968).

The branching option, which was called the Forced Signal Mode (FSM), was included to provide an automatic corrective learning sequence for S in the event of a missed signal. It forced S to receive four signals at the FIXED FAST signal rate, accompanied by KR and cueing. At the end of the timed

interval, the program automatically returned to the options which S had chosen prior to the FSM. The S could suppress the FSM at any time during the 6-min. TRAIN period except when FSM was activated.

The TEST mode presented the RANDOM SLOW signal rate for the entire 6-min. period with neither KR nor cueing available. At the end of the TEST period, summary knowledge of performance (detections, false alarms, and missed signals), was presented to S by means of the digital readouts. On the second day S monitored through three 6-min. periods to become acquainted with the equipment and task. On the third and fourth days, seven 6-min. periods were run. At the beginning of the third and fourth daily sessions, and at the end of the fourth session, each S was given summary knowledge of his cumulative performance in the TEST mode on mimeographed sheets.

RESULTS

The primary results consisted of the number of signal detections and the number of commissive errors (false alarms) produced by each S on the pretest and posttest days. The length of time spent in the various options by each S in Group T during the training days was also recorded.

Detections

Figure 1 shows the performance of Groups C and T on the pretest and posttest days with percentage of detections plotted against the four successive 12-min. time periods. A partially hierarchical analysis of variance was performed separately on the pretest and posttest detection data. The Ss were nested within groups, but common to the four time periods. The main effect of groups (G; training vs. control) and the Groups \times Periods interaction were not statistically significant at the .05 level, but the periods main effect was ($F = 3.48$; $df = 3, 96$). In the posttest analysis, the results indicated that the group effect was significant ($F = 8.40$; $df = 1, 32$; $p < .01$), while neither the Group \times Periods interaction nor the periods main effect was significant.

In order to test the relative gain of the two groups from the first to the final day, a similar analysis was performed on the total number of signals detected in each session (summed across periods) for each S on Days 1 and 5. The Ss' totals were nested within training conditions but common to the two

days, the pretest and posttest days being treated as time periods. The results of this analysis indicate that the days effect (D) was significant ($F = 40.17$; $df = 1, 32$; $p < .001$), as well as the Groups \times Days ($G \times D$) interaction ($F = 10.55$; $df = 1, 32$; $p < .01$). To test whether the increase shown by Group C was significant, a two-way analysis of variance was performed on their session totals, with the pretest and posttest days treated as time periods. A significant between-days effect ($p < .05$) was found.

Commissive Errors

The number of commissive errors in the pretest and posttest days for both groups is shown in Table 2. Poock and Wiener (1966) have pointed out that due to an extreme skewness in the distribution of commissive errors in this type of experiment, tests of goodness-of-fit always result in a significant deviation from the hypothetical uniform distribution. The median test diminishes skewness effects by counting the number of S s in each experimental group contributing more or less than the median number of commissive errors of the total sample. The contingency table thus formed can be analyzed using the chi-square statistic. The median number of commissive errors made in Days 1 and 5 was zero. In this extreme case, the contingency table was formed by counting the number of S s on each day making zero commissive errors, and those making one or more commissive errors. The test yielded a chi-square of 4.78, with $df = 1$. Even though the power of this test is reduced by the fact that S s were common to Days 1 and 5, this

TABLE 2
COMMISSIVE ERRORS FOR DAYS 1 AND
5 BY GROUPS

| Group | Day 1 | Day 5 | Total |
|----------|-------|-------|-------|
| Training | 33 | 5 | 38 |
| Control | 43 | 19 | 62 |
| Total | 76 | 24 | 100 |

value still indicates a significant ($p < .05$) decrease in the number of commissive errors from pretest to posttest. The median number of commissive errors on Day 5 was zero. A median test for groups on Day 5 was formed by again counting the number of S s in each group making zero commissive errors and those making one or more commissive errors. This yielded a chi-square less than unity, indicating no significant between-group difference in the contribution of commissive errors on Day 5.

Analysis of Training Options

Table 3 summarizes the percentage of time spent in the training options by each S as well as the percentage of signals received by each S , accompanied with KR or cueing. The percentage of signals include those received in the FSM. There was no significant correlation between individual option times and either the posttest data or the increase in performance from pretest to posttest. Table 3 shows that KR was called for 91.4% of the total TRAIN mode time, but cueing was relatively unpopular (35.5%). CUE, FIXED MEDIUM, and FIXED SLOW options dropped off sharply after the first training

TABLE 3
PERCENTAGE TIME IN TRAINING OPTIONS AND PERCENTAGE SIGNALS WITH KR
OR CUEING DURING DAYS 3 AND 4

| Training day | Percentage time | | | | | | | Percentage signals | |
|--------------|-----------------|--------------|------------|---------------|-------------|------|--------|--------------------|-------------|
| | Fixed fast | Fixed medium | Fixed slow | Random medium | Random slow | KR | Cueing | With Kr | With cueing |
| 2 | 23.5 | 24.2 | 7.9 | 21.5 | 22.9 | 82.5 | 40.3 | 84.8 | 54.2 |
| 3 | 23.6 | 13.3 | 7.7 | 19.2 | 36.2 | 93.0 | 32.0 | 93.8 | 43.4 |
| 4 | 27.0 | 9.7 | 8.4 | 23.1 | 31.8 | 95.0 | 36.5 | 94.0 | 47.4 |
| Total | 25.1 | 11.5 | 8.1 | 21.1 | 34.2 | 93.9 | 34.1 | 93.9 | 45.3 |

TABLE 4

TEST MODE CHOICE AND PERFORMANCE DATA
OVER THE THREE TRAINING DAYS

| Train- ing day | Percentage periods in test mode | Number of sig- nals pre- sented | Percent- age de- tections | Average commis- sive errors per test period |
|----------------------|---------------------------------------|--|---------------------------------|---|
| 2 | 35.3 | 87 | 86.2 | 0.555 |
| 3 | 47.0 | 272 | 93.3 | 0.196 |
| 4 | 54.6 | 328 | 94.2 | 0.062 |

session. Tables 3 and 4 point out the popularity of the two extreme signal rates (RANDOM SLOW and FIXED FAST options). The TEST period data are presented in Table 4. All attempts to correlate the performance in the TEST periods with the posttest performance or with *S* performance increase failed. The results from Table 4 show a steady increase in detection rates from Day 2 through Day 4, accompanied by an increase in the number of TEST periods chosen. At the same time, the average commissive error rate declined.

DISCUSSION

The results of this experiment show strong promise for the use of an autoinstructional approach to vigilance training. Superior signal detection performance of Group T over Group C in the final session, with no increase in commissive errors, indicates that an improvement in detection performance is not necessarily accompanied by an increase in the number of false alarms.

The significant (9.1%) increase in detections of Group C, from pretest to posttest, is the product of a steady increase over the five practice sessions. This result, as well as the "end-spurt" in the final 12-min. periods, is not uncommon in monitoring studies. Wiener (1968) found an increase of approximately 25% from pretest to posttest in a similar 5-day experiment. It is assumed that one reason for the improved performance of Group C was the added incentive introduced by informing *Ss* that their scores on the final day were most important to *E*. This argument might also be applied in some portion to Group T.

In the posttest, *Ss* in Group T showed very little variation in the percentage of detected signals over the four time periods, and varied only slightly from each other. This was due primarily to a "ceiling effect," the detection rates being so close to 100%. The homogeneous performance trend of Group T in the posttest explains the lack of correlation between use of options, which were highly variable, with the relatively nonvarying final performance data.

From the data presented in Table 3, one of the most noticeable features is the time spent in the KR option. The frequent choice of KR clearly points out that *Ss* sought knowledge of performance. In the NO-KR option, *Ss*' behavior was observed to be generally erratic, and *Ss* tended to play with the other options.

This experiment provides no evidence to support the training value of cueing. The *Ss*' choice of CUE was generally limited to the first half of a 6-min. period. The time spent in the CUE option generally dropped as the training progressed. This decrease in the use of cueing is possibly due to *Ss* becoming antagonistic toward the physical characteristics of the alerting tone. Another explanation is that once *S* learns (or relearns) the characteristics of the signal, he might desire more "challenge" from the task than the cueing option can provide. Wiener and Attwood (1968) have mentioned that *Ss* frequently volunteered the complaint that cueing denies them any challenge. The constant suppression of the FSM feature, and the apparent effort of *Ss* to evade the FSM when it was not suppressed, indicates that the importance of this corrective branching option lies not in what it teaches, but in its perception as a punishment rather than a training aid.

In conclusion, the autoinstructional approach to vigilance training results in increased signal detections without contributing to increased false responses. The erratic use of cueing during the training sessions suggests the reason for the failure of other experiments to produce consistently superior performance when the cueing techniques were rigidly applied. Annett and Paterson's (1966) conclusion, that the characteristics of the specific task should control the training pro-

cedure, is strengthened by this experiment. The suggestion of Swets et al. (1966) that S control his own training program also appears sound.

The increased performance attributed to this training method is apparently due to two factors. The first concerns simply the increased signal information content provided to S. The second stems from the positive effects that the training regime have on Ss when an interesting and self-participating training sequence is interjected between two monotonous vigilance sessions. The Ss readily volunteered the opinion that the training options made the task interesting and "fun." We have previously noted the "pin-ball machine effect," the self-motivating properties of devices which allow man to pit himself against a machine, with no opportunity for reward other than self-satisfaction and KR (Wiener, 1967). These factors should be carefully considered in future experimentation and practical use of autoinstructional trainers for perceptual tasks. Systems designers and training psychologists should be most eager to capitalize on what the trainee considers fun.

REFERENCES

- ADAMS, J. A., & HUMES, J. M. Monitoring of complex visual displays: IV. Training for vigilance. *Human Factors*, 1963, 5, 147-153.
- AIKEN, E. G., & LAU, A. W. Response prompting and response confirmation: A review of recent literature. *Psychological Bulletin*, 1967, 68, 330-341.
- ANNETT, J., & PATERSON, L. The use of cueing in training tasks: Phase II. (Technical Report No. 4119-1) Naval Training Device Center, 1966.
- ANNETT, J., & PATERSON, L. Training for auditory detection. *Acta Psychologica*, 1967, 27, 420-426.
- COLQUHOUN, W. P. Training for vigilance: a comparison of different techniques. *Human Factors*, 1966, 8, 7-12.
- HARDESTY, D., TRUMBO, D., & BEVAN, W. Influence of knowledge of results on performance in a monitoring task. *Perceptual and Motor Skills*, 1963, 16, 629-634.
- MACKWORTH, J. F. The effect of true and false knowledge of results on the detectability of signals in a vigilance task. *Canadian Journal of Psychology*, 1964, 18, 106-107.
- POOCK, G. K., & WIENER, E. L. Music and other auditory backgrounds during visual monitoring. *Journal of Industrial Engineering*, 1966, 17, 318-323.
- SWETS, J. A., MILLMAN, S. H., FLETCHER, W. E., & GREEN, D. M. Learning to identify nonverbal sounds: An application of a computer as a teaching machine. *Journal of the Acoustical Society of America*, 1962, 34, 928-935.
- SWETS, J. A., HARRIS, J. R., MCELROY, L. S., & RUDLOE, H. Computer-aided instruction in perceptual identification. *Behavioral Science*, 1966, 11, 98-104.
- WEISZ, A. A., & MCELROY, L. S. Response and feedback techniques for automated training of visual identification skills. (Technical Report No. 789-3) Naval Training Device Center, 1964.
- WIENER, E. L. Knowledge of results and signal rate in monitoring: A transfer of training approach. *Journal of Applied Psychology*, 1963, 47, 214-222.
- WIENER, E. L. Can vigilance be taught? *Human Factors in Electronics. Technical Digest 28C72*. New York: Institute of Electrical and Electronics Engineers, 1967. (Abstract)
- WIENER, E. L. Training for vigilance: Repeated sessions with knowledge of results. *Ergonomics*, 1968, 11, 547-556.
- WIENER, E. L., & ATTWOOD, D. A. Training for vigilance: Combined cueing and KR. *Journal of Applied Psychology*, 1968, 52, 474-479.

(Received April 25, 1968)

KNOWLEDGE OF PERFORMANCE AS AN INCENTIVE IN REPETITIVE INDUSTRIAL WORK

P. S. HUNDAL¹

Panjab University, Chandigarh-14

This study was designed to assess the purely motivational effects of knowledge of performance in a repetitive industrial task. The Ss were low paid workers with a few years (1-5 yr.) experience on the job. The experimental task was to grind a metallic piece to a specified size and shape. Experimental conditions were imposed a week before starting the experiment. The workers adjusted readily since the experimental conditions did not interfere with the work. Eighteen male workers were divided randomly into three groups. The Ss in Group A received no information about their output; Ss in Group B were allowed a rough estimate of their output; Ss in Group C were given accurate information about their output and could check it further by referring to a figure displayed before them. Results show increased output with increases in degree of knowledge of performance.

A review of studies of the effects of knowledge of results on performance by Ammons (1954) suggests that knowledge of results (KR), universally, tends to improve the performance of Ss in laboratory situations. It is difficult, however, to know from most studies whether the improvement is due to motivational effects of knowledge of results or to some side effect such as "information" or "reward" which has not been controlled systematically.

Gibbs and Brown (1955), for the first time, tried to isolate and measure the motivational aspects of KR by designing an experiment so that KR was more casual and accidental than is usually the case. Under these conditions, they argued, the increase in output, if any, can be attributed to purely motivational impact of KR. In their study they found significant improvements in performance of S as a function of KR.

In a replication of the above study, with groups having different degrees of KR, Chapanis (1964) failed to confirm the Gibbs and Brown findings. Quite recently Locke and Bryan (1966, 1967) have reported that there is no effect of knowledge of scores on performance. Their studies show that performance increases when Ss adopt goal setting procedures; that is, high goals lead to higher levels of performance than low goals.

Further, very little is known about the purely motivational impact of KR on the performance of industrial workers in their actual work situation.

The present study is a step in this direction. The experiment was arranged on the lines of Chapanis (1964). Industrial workers acted as Ss and their performance on the job was manipulated in such a way that it was possible to control differential, casually given KR. It was hoped that under such conditions each S's output would be determined entirely by self-competition and his own satisfaction in working.

METHOD

Subjects

The Ss were 18 male industrial workers employed in a small industrial unit. They had been with this factory for 1 to 5 yr. Their age ranged from 24 to 37 yr. They belonged to the lower income group, their salary ranging from 150 rupees to 175 rupees/mo. Their work was supervised by an assistant manager, a partner in the concern. The Ss' task was to "grind" a metallic piece to a specified size and shape. Finished pieces were placed in a nearby box.

Procedures

The 18 Ss were assigned randomly to one of three groups, 6 Ss to a group.

Group A. The Ss in this group were required to keep their finished pieces in boxes fitted with flaps so they could not see how many they had done. Their boxes were emptied during the lunch break and at the end of the workday.

¹ Requests for reprints should be sent to the author, Department of Psychology, Panjab University, Chandigarh-14, India.

TABLE 1
AVERAGE OUTPUT AND THE RANK OF THE PREEXPERIMENTAL
AND EXPERIMENTAL PERIODS

| Preexperimental period | | | | | | Experimental period | | | | | |
|------------------------|------|---------|------|---------|------|---------------------|------|---------|------|---------|------|
| Group A | | Group B | | Group C | | Group A | | Group B | | Group C | |
| Output | Rank | Output | Rank | Output | Rank | Output | Rank | Output | Rank | Output | Rank |
| 38 | 6.0 | 41 | 17.0 | 40 | 13.0 | 40 | 5.5 | 38 | 2.5 | 48 | 18.0 |
| 40 | 13.0 | 39 | 9.0 | 34 | 1.0 | 35 | 1.0 | 40 | 5.5 | 40 | 5.5 |
| 38 | 6.0 | 41 | 17.0 | 40 | 13.0 | 38 | 2.5 | 47 | 17.0 | 45 | 15.0 |
| 36 | 3.0 | 38 | 6.0 | 39 | 9.0 | 43 | 10.5 | 44 | 13.0 | 43 | 10.5 |
| 39 | 9.0 | 35 | 2.0 | 37 | 4.0 | 44 | 13.0 | 40 | 5.5 | 46 | 16.0 |
| 41 | 17.0 | 40 | 13.0 | 40 | 13.0 | 41 | 8.0 | 42 | 9.0 | 44 | 13.0 |

Note.—Mean output for Preexperimental period, Group A, is 38.7; Group B, 39.0; Group C, 38.3. Mean output for Experimental period, Group A, is 40.2; Group B, 41.8; Group C, 44.3.

Group B. Storage boxes for these Ss had no flaps. The Ss could see their mounting piles of finished pieces. The periodic cleaning of their boxes was always partial; thus, there always was a pile of finished pieces in their boxes at the beginning of each work session. The workers could, therefore, make a rough estimate of their output.

Group C. The Ss in this group could also see their processed pieces in their respective boxes. After each work session their boxes were completely emptied. They could, therefore, make better estimates of their total output than workers in either of the other two groups. Moreover, they could also see a number, displayed on a card beside the box, that was an index of the amount of work completed during the preceding session. The Ss were not, however, told what the number was. They could either draw an inference from it or ignore it.

All the persons worked under the same roof. The Ss in different groups were put together by shifting their work seats. The experiment started 1 wk. after they had worked in their new places under usual factory conditions. This week, the preexperimental period, was used to gather production records before undertaking the KR manipulation. A fixed quantity of raw material was supplied to all Ss at regular intervals. By keeping them constantly overloaded with raw material, they could not estimate how much work they had done by referring to the stock of raw material.

The average output of all three groups was comparable before undertaking the experiment.

The experimental manipulation began on May 22, 1967, and ended on the 27th of the same month. The entire arrangement was done by the supervisor. The *E* did not come in direct contact with *S*; he was, however, available for consultation.

The hypothesis under study was that the output of these Ss should increase under the impact of subtle variations in knowledge of performance.

RESULTS

The basic data gathered are the average number of pieces processed by each *S* in the preexperimental and the experimental periods. The significance of differences in output between the three groups at both stages was checked by Kruskal-Wallis' one-way analysis of variance (Siegel, 1956).

The average output for each *S* and for each group is reported in Table 1 separately for the preexperimental and experimental periods. All Ss under these two conditions were ranked separately on the basis of their output. The rank values, given in Table 1, were then used to compute *H* values as detailed by Siegel (1956, Formula 8.1). The *H* values for the preexperimental and experimental groups worked to be .28 and 4.17, respectively. The latter value is significant at the .07 level (one-tailed test).

Second, mean output of each of the experimental groups was compared with that of its counterpart in the preexperimental period. The main objective was to examine how group performance during the experimental period compared with earlier performance during the preexperimental period. Significance of differences was estimated by means of the Wilcoxon matched-pair sign rank test (Siegel, 1956, pp. 77-79). Three comparisons were made, one each for Groups A, B, and C. Differences for Groups A and B were not sig-

nificant. The difference for Group C was significant at the .025 level (one-tailed test).

DISCUSSION

A look at Table 1 reveals that the mean output of the groups increases in the experimental period in direct relation to their degree of awareness of their performance. The difference approaches statistical significance ($p < .07$, one-tailed test).

The second comparison involving intra-group differences between preexperimental and experimental periods shows the greatest (and statistically significant) increases in the output of Group C. The corresponding increases in Group A and B are not significant. Both approaches to the analysis of results agree in showing that output tends to increase under the presumed impact of the purely motivational aspects of KR.

REFERENCES

- AMMONS, R. S. *Knowledge of performance: Survey of literature, some possible applications and suggested examination*. (Technical Report No. 54-14) Washington, D. C.: United States Air Force, 1954.
- CHAPANIS, A. Knowledge of performance as an incentive in repetitive monotonous tasks. *Journal of Applied Psychology*, 1964, **48**, 263-267.
- GIBBS, C. B., & BROWN, I. D. *Increased production from the information incentive in a repetitive task*. (Technical Report No. 230) Great Britain: Medical Research Council, Applied Psychological Research Unit, 1955.
- LOCKE, E. A., & BRYAN, J. F. Cognitive aspects of psychomotor performance: The effects of performance goals on level of performance. *Journal of Applied Psychology*, 1966, **50**, 286-291.
- LOCKE, E. A., & BRYAN, J. F. Performance goals as determinants of level of performance and boredom. *Journal of Applied Psychology*, 1967, **51**, 120-130.
- SIEGEL, S. *Nonparametric statistics*. New York: McGraw Hill, 1956.

(Received April 26, 1968)

WORK SHIFT, OCCUPATIONAL STATUS, AND THE PERCEPTION OF JOB PRESTIGE¹

RONALD H. BOHR² AND ARNOLD B. SWERTLOFF³

Philadelphia State Hospital

Ratings of job prestige provided by 170 respondents in two widely different work settings, a state mental hospital and an electronics plant, revealed a significant interaction between work shift and occupational status level. As hypothesized, nonsupervisory night workers (hospital attendants and assembly line workers) attributed less prestige to own and co-workers' jobs than did their counterparts on the day shift, whereas no differences were found between the ratings of day and night supervisory personnel (registered nurses, industrial engineers, and foremen). These results suggest that situational characteristics of the work environment are more potent determinants of the job attitudes of lower-echelon than of upper-echelon workers.

It is surprising that while work shift has been found in several investigations to be closely related to a wide range of social, psychological, and physiological variables (Mott, Mann, McLouglin, & Warwick, 1965, Ch. 1), it is generally ignored in studies of job attitudes. This study presents data gathered from day- and night-shift workers on one significant attitude, namely, the perception which workers have of the prestige of their own and co-workers' jobs.

Specifically, it is hypothesized that prestige ratings are related to an interaction between work shift and the status of the rater, and that shift and prestige ratings are related for low status (nonsupervisory) workers, with night workers attributing less status to jobs than day workers, but not for high status (supervisory) workers regardless of shift. It is theorized that while the job perceptions of low status workers are closely related to certain situational characteristics of their work setting (like their hours of work), those of high status workers are relatively self-directed and impervious to such situational determinants. Hence, since night work is generally less prestigious than day work, low status night workers might feel that all night

workers hold jobs of low prestige. Conversely, if the perceptions of high status night employees are determined by internalized referents (e.g., a "cosmopolitan" professional identity), they will not differ from their counterparts on the day shift.

METHOD

This hypothesis was tested in two widely different work settings, a state mental hospital and an electronics plant, in order to give some indication of the generalizability of the findings.⁴ Occupational status in both settings was defined as the occupancy of a supervisory or a nonsupervisory job. In the hospital, which operated on a three-shift schedule, day workers were contrasted with both evening and night workers. The electronic plant had only two shifts. In all, 170 workers were employed as Ss.

Subjects

State mental hospital. Of the 234 nursing service personnel at a 600-bed hospital, 96 completed the ratings of job prestige. Of these 96 respondents, 24 were nurses (65% of all nurses) and 72 were attendants (37% of all attendants). Half of each group worked on the day shift and the other half on either the evening or night shift.

Electronics plant. A random stratified sample of 74 engineers, foremen, and assembly line workers was selected from a total of over 3,000 plant employees. The day shift sample included 23 engineers and foremen (supervisory personnel) and 23 assembly line workers (nonsupervisory personnel). The night shift sample included 14 employees in each of the two categories.

¹ An earlier version of this paper was reported at the Eastern Psychological Association, Washington, D. C., April 1968.

² Requests for reprints should be sent to Ronald H. Bohr, Coordinator, Psychosocial Research, Rehabilitation Center, Building S-9, Philadelphia, Pennsylvania 19114.

³ Now at the Internal Revenue Service, Washington, D. C.

⁴ Thanks are extended to Aaron Smith of Haverford State Hospital, Haverford, Pa., and to Albert Lagore of the Ford-Philco plant, Philadelphia, Pa., for their assistance in obtaining respondents.

TABLE 1
ANALYSIS OF VARIANCE OF THE PRESTIGE RATINGS
OF MENTAL HOSPITAL NURSING PERSONNEL

| Source | df | MS | F |
|----------------|----|---------|----------|
| Shift (A) | 1 | 829.00 | 11.95* |
| Raters (B) | 1 | 236.00 | |
| Jobs rated (C) | 1 | 4116.00 | 101.25** |
| A × B | 1 | 360.00 | 5.19* |
| A × C | 1 | 11.00 | |
| B × C | 1 | 94.00 | |
| A × B × C | 1 | 1.00 | |

* $p < .05$.
** $p < .01$.

Procedure

The instrument used to determine the prestige of various occupations, a modified form of Mishler and Tropp's (1956) status scale, was composed of a list of 14 community jobs (with corresponding numerical values) arranged in descending order from Supreme Court Justice (100) to Garbage Collector (50). Respondents indicated for each of several hospital or factory jobs the community job most equivalent in prestige. Hospital personnel rated the prestige of Nurse and Attendant; plant employees rated the prestige of Industrial Engineer, Foreman, and Assembly Worker.

RESULTS

Separate analyses of variance were performed on the ratings made by hospital and industrial workers. As Table 1 indicates, the $2 \times 2 \times 2$ (Shift \times Raters \times Jobs Being Rated) analysis of the hospital data revealed a significant interaction between shift and rater, $F(1, 92) = 5.19, p < .05$. Overall mean differences between shifts were computed by subtracting night shift means from day shift means. Overall mean differences between day and night shift for nurses and attendants were $- .60$ and 5.6 , respectively.

Table 2 shows the $2 \times 2 \times 3$ (Shift \times Raters \times Jobs Being Rated) analysis of variance for the industrial sample. A significant interaction between shift and rater was obtained, $F(1, 70) = 5.08, p < .05$. Overall mean differences between day and night shift workers were -1.00 for engineers and foremen and 3.60 for assemblers. Since the ratings of both types of supervisory personnel (engineers and foremen) were combined, Table 2 shows one degree of freedom for raters (B) but two for the three jobs rated (C).

In terms of the specific community jobs

seen as equivalent, all hospital nurses felt they had slightly more prestige than public school teachers ($M[\text{day}] = 87.4; M[\text{night}] = 88.1$), and that attendants were similar to typists ($M[\text{day}] = 75.8; M[\text{night}] = 76.3$). Among attendants, however, day workers considered themselves more prestigious than electricians ($M = 83.4$), while night workers rated themselves just above typists ($M = 77.1$). Similarly, while day attendants thought nurses resembled airline pilots ($M = 91.3$), night attendants rated them with teachers ($M = 86.1$). In the factory, all supervisory personnel thought themselves similar to teachers ($M[\text{day}] = 85.3; M[\text{night}] = 86.1$), and considered assemblers comparable to semi-skilled workers in an auto factory ($M[\text{day}] = 67.7; M[\text{night}] = 68.7$). However, while night assemblers concurred in this judgment ($M = 67.0$), day assemblers felt they were more like store clerks ($M = 70.4$). Also, while day assemblers viewed supervisors as just below teachers in prestige ($M = 83.6$), night assemblers rated them equal to electricians ($M = 80.0$).

DISCUSSION

The present results support the hypothesis that prestige ratings of low status workers are related to their hours of work, while those of high status workers are not. In light of the concern expressed about shift work (Mott et al., 1965; Pearlin, 1962), it seems noteworthy to illustrate one way in which it is differentially related to job attitudes.

Although further research is needed, it is suggested here that the present results reflect a difference in the frame of reference of vari-

TABLE 2
ANALYSIS OF VARIANCE OF THE PRESTIGE RATINGS
OF INDUSTRIAL EMPLOYEES

| Source | df | MS | F |
|----------------|----|---------|---------|
| Shift (A) | 1 | 87.00 | 190.24* |
| Raters (B) | 1 | 206.00 | |
| Jobs rated (C) | 2 | 6375.00 | 5.08* |
| A × B | 1 | 262.00 | |
| A × C | 2 | 0.00 | |
| B × C | 2 | 322.00 | 9.61** |

* $p < .05$.
** $p < .01$.

ous workers. Indeed, a previous study (Bohr & Goldman, 1967) demonstrated that prestige ratings were related to the nature of the immediate work setting for attendants but not for upper status mental hospital employees. Both investigations indicate that situational characteristics of the work environment are more potent determinants of the job attitudes of lower-echelon than of upper-echelon workers.

REFERENCES

- BOHR, R. H., & GOLDMAN, A. R. Changing status of mental health occupations in a transitional hospital. *Proceedings of the 75th Annual Convention of the American Psychological Association*, 1967, 2, 233-234.
- MISHLER, E. G., & TROPP, A. Status and interaction in a psychiatric hospital. *Human Relations*, 1956, 9, 187-205.
- MOTT, P. E., MANN, F. C., McLOUGHLIN, Q., & WARWICK, D. P. *Shift work: The social, psychological, and physical consequences*. Ann Arbor: University of Michigan Press, 1965.
- PEARLIN, L. I. Alienation from work: A study of nursing personnel. *American Sociological Review*, 1962, 27, 314-326.

(Received April 27, 1968)

COMPARATIVE STUDY OF NEED SATISFACTIONS IN GOVERNMENTAL AND BUSINESS HIERARCHIES

J. B. RHINEHART,¹ R. P. BARRELL,² A. S. DEWOLFE,²
J. E. GRIFFIN,³ AND F. E. SPANER⁴

Veterans Administration Hospital, Coatesville, Pennsylvania

Two thousand twenty-six questionnaires from managers (supervisors) of a government agency, Veterans Administration, Department of Medicine and Surgery (VA-DM&S), were evaluated and compared with Porter's (1962) Business and Industry (B&I) sample. Satisfaction decreased from top- to lower-management levels and the greatest satisfaction deficit at all levels was in autonomy and self-actualization for both DM&S and B&I Ss. Dissatisfaction for DM&S was markedly greater than for B&I, confirming the study of Paine, Carroll, and Leete (1966) who found 95 government managers less satisfied than B&I Ss. Government's lag in the human relations area contrasted with B&I's growing people-centered orientation is offered as a possible explanation for the need-satisfaction differences between the two groups.

The dimensions of the personnel problem of the federal government have been of considerable concern for the past decade. David T. Stanley (1964) wrote:

In the foreseeable future the federal government will continue to compete with commerce and industry for personnel possessing skills that are scarce and will continue to be scarce. . . . The government's competitive position will be better than it was because of the 1964 salary increases and because the market . . . is easing a little as government procurement is cut down. Nevertheless, trend analyses generally show long-term shortages. . . . At present it is uncertain whether the educational system can meet these shortages. They will be mitigated, but not really solved, by in-service training, outside training assignments, and personnel utilization improvements [p. 16].

Stanley's perception of the problem is not an isolated one. Numerous committee reports and such publications as *Executives for Government* (David & Pollock, 1957), *The American Federal Executive* (Warner, Van Riper, Martin, & Collins, 1963), and *The Job of the Federal Executive* (Bernstein, 1958) identify and explain the government's strengths and weaknesses as it competes for talented manpower in the 1960s.

¹ Requests for reprints should be sent to Jesse B. Rhinehart, Psychology Service, Veterans Administration Hospital, Coatesville, Pennsylvania 19320.

² At Veterans Administration Hospital, Downey, Illinois.

³ Now at Veterans Administration Hospital, Hines, Illinois.

⁴ Now with Department of Health, Education and Welfare, Chevy Chase, Maryland.

In 1959, Kilpatrick, Cummings, and Jennings (1964) initiated a unique approach to the personnel crisis in government. In a large-scale survey in which personal interviews were conducted with more than 5,000 respondents on standardized questionnaires, the researchers attempted to (a) analyze the occupational values and attitudes toward work that prevail in American society today and (b) ascertain the attitudes of various groups in the American public toward the federal civilian service generally and toward the federal service as an employer. The most significant single conclusion emerging from the study was this: The image of federal employment is markedly out of phase with the occupational values of those whom the service most needs to attract to its ranks.

Shortages of manpower in key occupations of the government service, plus the indifferent image of federal employment noted in the preceding research citation, served to motivate a second major research project, the results of which were published under the title *The Higher Civil Service* (Stanley, 1964). Originally proposed as an evaluation of "rank-in-job" versus "rank-in-man" principles, this study of 16,000 federal executives broadened into a study of policies, procedures, and institutional concepts for the purpose of achieving a realistic assessment of the government's use of higher scientific, professional, and managerial employees. In one segment of the study, a small number (80) of former federal

employees were asked why they had resigned. When the results were analyzed, it was found that 34% reported leaving for new opportunities and new kinds of work experiences and 20% reported leaving because of dissatisfaction with programs, policies, colleagues, and frustrations. It can be assumed, in short, that more than one-half of these former employees left their government positions because of failure to find their jobs satisfying. Herein may reside the core of the federal manpower problem.

Porter (1962) employed a questionnaire utilizing a modified Maslow-type (1954) categorization of needs to determine how managers in business and industry feel about their jobs. Subsequently, similar investigations have been extended to the military (Porter and Mitchell, 1967) and to union officials (Miller, 1966). To date, only one study (Paine, Carroll, & Leete, 1966) using Porter's technique has assessed the job satisfaction of civil service personnel. When the job satisfactions of 95 managers in a government agency were compared with the job satisfactions of Porter's group from business and industry, the former evidenced less satisfaction in all need items included in the questionnaire.

The present study, the first of six projected, proposes to determine how managers (supervisors) of the Department of Medicine and Surgery of the Veterans Administration compare with their counterparts in business and industry (Porter, 1962) in terms of the extent to which five need areas (Security, Social, Esteem, Autonomy, Self-Actualization) are met by their jobs. The resultant information, it is believed, can serve ultimately to reduce the manpower problem of one segment of a federal agency by providing a base for program innovations which will (a) decelerate personnel attrition through resignations and (b) increase the job performance and satisfactions of managers, with concomitant improvement in the productivity of their subordinates.

METHOD

Procedure and Sample

Porter's (1961) questionnaire, the data sheet of which was revised to fit the agency position structure more appropriately, was distributed to super-

visory personnel in 148 hospitals, domiciliaries, and outpatient clinics through personnel officials designated as coordinators for this project. The number of questionnaires mailed was based upon estimates provided by each coordinator of the number of supervisory personnel at his station eligible to participate in the study. Those eligible to participate were supervisors drawn from the three classes of positions in the Department of Medicine and Surgery (DM&S) of the agency: (a) Title 38, including physicians, dentists, and nurses in the medical professions, (b) General Schedule, comprising professional, semi-professional, administrative, technical and clerical personnel, nonmedical professional and white collar positions, and (c) Wage Administration Supervisors, consisting of skilled and unskilled employees in trades, crafts, maintenance, and housekeeping.

To elicit cooperation and assure anonymity for each participant, the coordinators were asked to assemble the potential participants in groups as large as were feasible, to explain to them the nature of the study and its potential value, and to request their cooperation. Further, they were to provide an opportunity to answer any questions about the study, and to distribute packets of materials. The individual packets included the questionnaire and a letter from the principal investigator with instructions for completing the questionnaire and directions for mailing it in an attached, franked envelope addressed to the Psychology Department of a cooperating university where the envelope was to be opened and destroyed to eliminate postmark identification. The questionnaires were then mailed, in bulk, to the principal investigator.

This procedure, geared to permissive, anonymous participation, resulted in a return of 9,841 questionnaires of 16,293 mailed to the local coordinator, or 60%. For replication purposes, however, only the first 2,026 questionnaires were employed for comparison with Porter's (1962) sample of 1,916 from business and industry. All of the completed questionnaires will be utilized for future, extensive, intra-agency studies.

For the 2,026 questionnaires constituting the DM&S data of this study, subdivision into four management levels seemed most appropriate. Therefore, for comparison purposes, it was necessary to convert Porter's five management levels to four by combining his two top levels into one, designated here as Top Management and consisting of the 114 cases in his President category (19%) and the 611 cases in his Vice-president category (81%). The resulting organization of the two groups of data into four management levels is shown in Table 1.

Some characteristics of the B&I data and DM&S data and the distribution of totals by four management levels and by four age levels is shown in Table 2.

Comparison of the two groups indicates that DM&S respondents are older at all management levels, and, except for Top Management, have lower educational levels.

TABLE 1

RESULTS

COMPARISON OF THE FOUR MANAGEMENT LEVELS IN THE BUSINESS AND INDUSTRY AND DEPARTMENT OF MEDICINE AND SURGERY DATA

| Management levels | Business and industry | Department of medicine and surgery |
|-------------------|---|---|
| Top | Presidents Vice-presidents | Directors Assistant directors Chiefs of staff |
| Upper-middle | Division managers Plant managers Department managers | Chiefs Assistant chiefs of divisions and services |
| Lower-middle | Approximate level of department and sub-department managers | All between lower- and upper-middle, primarily section chiefs |
| Lower | First- or second-level supervisors | Unit chiefs |

The Questionnaire and Its Scoring

Only one segment of Porter's questionnaire, the rationale of which has been described elsewhere (Porter, 1961) provided data for this study. The results reported here are based on responses to 13 items relevant to a Maslow-type classification and organized according to their prepotency into five types of needs: security, social, esteem, autonomy, and self-actualization. For each of the 13 specific items (e.g., "the feeling of esteem") respondents were asked to indicate on a 1-7 rating scale: (a) How much is there now? (b) How much should there be?

The amount of need satisfaction experienced in his management position by each DM&S respondent for each of the 13 items was determined by subtracting his response to Part *a* of the item from his response to Part *b* of the item. Individual scores were then averaged on all 13 items for each of the four management levels and reported in terms of the five need categories. Either a sign test or signed-rank test was used to test for statistical significance (Siegel, 1956).

Table 3 compares the need satisfaction of B&I supervisors or managers on the basis of level of position. It should be noted that, following the example set by Paine, Carroll, and Leete (1966), in Table 3 the term "average satisfaction" is substituted for Porter's "perceived deficiencies in need fulfillment" with high numbers indicating less satisfaction.

Porter (1962) concluded that the more satisfied managers tended to cluster at the highest management level and that satisfaction tended to decrease at each successive lower level of management. When this conclusion is evaluated on the basis of Porter's data as presented in Table 3, the projected pattern of decreasing need satisfaction is apparent; a similar pattern prevails for the DM&S data. The differences constituting the patterns are significant at the .01 level by signed-rank tests.

From Table 3 it may be observed that DM&S managers are consistently more dissatisfied than B&I managers at the top three management levels. These differences were found to be statistically significant at the .01 level for the top and upper-middle levels and at the .001 level for the lower-middle groups. When the two lower management levels are compared, this greater dissatisfaction for DM&S is not demonstrated. It is at this point in the management hierarchy, it should be noted, that the demographic data reveal the greatest disparity in educational backgrounds with 76.2% and 18%, respectively,

TABLE 2
COMPARISON OF DM&S WITH B&I IN TERMS OF DISTRIBUTION OF *N* OF TOTAL SAMPLE BY FOUR MANAGEMENT LEVELS AND FOUR AGE GROUPS, AND CHARACTERISTICS OF SAMPLE BY MANAGEMENT LEVEL

| Management level | Age group | | | | | | | | Total <i>N</i> for level | | Median age | | College degree (%) | |
|------------------|-----------|------|-------|------|-------|------|-----|------|--------------------------|------|------------|------|--------------------|------|
| | 20-34 | | 35-44 | | 45-54 | | 55+ | | | | | | | |
| | B&I | DM&S | B&I | DM&S | B&I | DM&S | B&I | DM&S | B&I | DM&S | B&I | DM&S | B&I | DM&S |
| Top | 59 | 0 | 271 | 11 | 263 | 58 | 132 | 59 | 725 | 128 | 46.0 | 54.2 | 74.0 | 85.0 |
| Upper-middle | 95 | 33 | 288 | 165 | 206 | 298 | 70 | 172 | 659 | 668 | 43.1 | 49.6 | 75.0 | 71.0 |
| Lower-middle | 100 | 43 | 208 | 173 | 98 | 281 | 25 | 94 | 431 | 591 | 40.6 | 47.8 | 75.0 | 56.0 |
| Lower | 32 | 57 | 46 | 201 | 14 | 308 | 9 | 73 | 101 | 639 | 39.2 | 47.0 | 76.2 | 18.0 |

TABLE 3

COMPARISON OF B&I AND DM&S IN TERMS OF AVERAGE NEED SATISFACTION SCORES
FOR EACH NEED CATEGORY AND FOUR MANAGEMENT LEVELS

| Need category | Management level | | | | | | | |
|----------------------|------------------|------|--------------|------|--------------|------|-------|------|
| | Top | | Upper-middle | | Lower-middle | | Lower | |
| | B&I | DM&S | B&I | DM&S | B&I | DM&S | B&I | DM&S |
| Security | .40 | .63 | .40 | .57 | .38 | .56 | .80 | 1.32 |
| Social | .30 | .53 | .33 | .48 | .33 | .51 | .61 | .70 |
| Esteem | .43 | .55 | .66 | .82 | .71 | .88 | 1.21 | 1.04 |
| Autonomy | .50 | .84 | .87 | .98 | .95 | 1.24 | 1.47 | 1.32 |
| Self-actualization | .87 | .99 | 1.12 | 1.26 | 1.17 | 1.47 | 1.70 | 1.53 |
| <i>M</i> of 13 items | .53 | .74 | .76 | .90 | .81 | 1.04 | 1.28 | 1.21 |

Note.—The larger the number, the less the need satisfaction.

for the B&I and DM&S lower level supervisors. This suggests the possibility that the greater dissatisfaction of these B&I managers in comparison with their DM&S counterparts may be a function of their higher educational level which could be indicative of higher levels of aspiration.

When both population samples were evaluated to determine their relative average satisfaction, rankings (from most to least deficient) for the five need areas revealed certain similarities. For both groups Self-Actualization and Autonomy ranked first and second; Social ranked fifth. However, B&I ranked Esteem third with Security fourth; the reverse ranking occurred for DM&S.

Though the patterns of relative perceived need deficiencies for both groups were similar, in all of the five need areas and for all management areas, the degree of dissatisfaction was significantly greater for DM&S ($p < .01$). When assessed in terms of the five individual need categories, the greater dissatisfaction of DM&S supervisors was found to be statistically significant for the Social and Self-Actualization needs, both at the .01 level. When the comparison is confined to the three top management levels, for both Self-Actualization and Autonomy the differences between B&I and DM&S are significant at the .05 level. For all other category comparisons for all four levels of management as well as for the top three levels, differences between B&I and DM&S were in the direction of

greater dissatisfaction for DM&S supervisors, but these differences were not found to be statistically significant.

In a comparative study of need satisfactions in military and business hierarchies, Porter and Mitchell (1967), using three management levels, concluded that the influence of military rank on need satisfaction may be greater than the effect of echelon level in business organizations. When the ranges of average satisfaction (for the total of all 13 items) from the highest to the lowest of the top three B&I and DM&S management levels are observed in Table 3, it is apparent that the range for DM&S (.74–1.04) closely approximates that for B&I (.53–.81), suggesting that the effect of echelon level is about the same for both these governmental and nongovernmental supervisors.

CONCLUSIONS

When managers (supervisors) from the Department of Medicine and Surgery of the Veterans Administration are compared with managers from business and industry, several conclusions are apparent:

First, both groups show positive relationships between vertical location in the management hierarchy and need satisfaction, with satisfaction decreasing as the management scale is descended.

Second, for both groups the two highest order needs, Autonomy and Self-Actualization, are less well satisfied at all management

levels. However, for the government supervisors the Security need is less well satisfied than Esteem, while the reverse is true for the managers from business and industry.

Third, although the position-level-satisfaction patterns and need deficits were similar, the degree of dissatisfaction of the government group was markedly greater when assessed in terms of the over-all needs and all four management levels. When the top three groups are compared level-to-level, the DM&S supervisors reveal more dissatisfaction for each management segment. However, this finding does not hold for the lower level supervisors.

Fourth, in contrast to the military, apparently the effect of echelon level on satisfaction is about the same for both the governmental and nongovernmental groups studied here.

Fifth, the percentage of college graduates was higher for the business managers at the three lowest levels but higher for the government managers at the highest level, suggesting that a college degree is a greater requisite for reaching the top in government than in business.

Sixth, at each level, the government managers were older than their counterparts in business. Whether this means that employees entered government service at a later age or required longer service before promotion to a supervisory role materialized could not be determined.

In this and Porter's study, it will be recalled that job satisfaction was quantified by subtracting the participants' assessment of "reality" (How much is there now?) from his "expectation" (How much should there be?). Dissatisfaction, therefore, is, in part, a function of level of expectation. In short, the norms for government managers and business managers could be different, as pointed out by Paine, Carroll, and Leete (1966) in their study, the findings of which "... indicated that government agency managers had much less need satisfaction than private industry managers similar to them in age and organizational level [p. 249]." These writers also suggested that studies of other government agencies would be needed to confirm their findings. The present study, employing

a much larger number of respondents (2,026 as compared to 95), certainly seems to have produced sufficient confirmation of their finding.

The answer to the question of why government managers are less satisfied than managers in private business is doubtless so complex that many aspects of the work climate must be taken into account. However, one possible facet may reside in the failure of government to emulate industry's increasing concern with human relations factors in the work situation, a concern initiated by the famed Hawthorne Studies at the Western Electric plant in Chicago. These studies, completed in 1927, represented the first honest and concerted effort to understand employees, instead of approaching the problem solely from the managerial point of view of improving efficiency on an economic level. For the past 35-40 yr., industry has made increasing use of the behavioral sciences in its slow but definite movement from rule-centered and work-centered orientations to people-centered and group-centered orientations. Some substantiation of government's lag in the human relations area is to be found in a 1968 unpublished report of a questionnaire survey to which 108 federal executives responded, indicating their preferences for various seminar topics:

The results seem to indicate that, on the whole, respondents disliked topics and general areas associated with the behavioral sciences and with the psychological growth of the executive (Rhinehart, 1968).

Perhaps an increased concern for human relations could contribute to increasing satisfaction among government managers. It would seem logical that increased manager-satisfaction would result in increased satisfaction among those they supervise.

REFERENCES

- BERNSTEIN, M. H. *The job of the federal executive*. The Brookings Institution, Washington, D. C., 1958.
- DAVID, P. T., & POLLOCK, R. *Executives for government*. The Brookings Institution, Washington, D. C., 1957.
- KILPATRICK, F. P., CUMMINGS, M. C., JR., & JENNINGS, M. K. *The image of the federal service*. The Brookings Institution, Washington, D. C., 1964.

- MASLOW, A. H. *Motivation and personality*. New York: Harper, 1954.
- MILLER, E. L. Job attitudes of national union officials: Perceptions of the importance of certain personality traits as a function of job level and union organizational structure. *Personnel Psychology*, 1966, 4, 395-410.
- PAINE, F. T., CARROLL, S. J., JR., & LEETE, B. A. Need satisfactions of managerial level personnel in a government agency. *Journal of Applied Psychology*, 1966, 3, 247-249.
- PORTER, L. W. A study of perceived need satisfactions in bottom and middle management jobs. *Journal of Applied Psychology*, 1961, 45, 1-10.
- PORTER, L. W. Job attitudes in management: Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology*, 1962, 46, 375-384.
- PORTER, L. W., & MITCHELL, V. F. Comparative study of need satisfactions in military and business hierarchies. *Journal of Applied Psychology*, 1967, 51, 139-144.
- RHINEHART, J. B. Results of the F. E. B. Seminar Questionnaire: 1968 (Unpublished).
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- STANLEY, D. T. *The higher civil service*. The Brookings Institution, Washington, D. C., 1964.
- WARNER, W. L., VAN RIPER, P. P., MARTIN, N. H., & COLLINS, O. F. *The American federal executive*. New Haven, Conn.: Yale University Press, 1963.

(Received April 29, 1968)

BEYOND PARKINSON'S LAW:

III. THE EFFECT OF PROTRACTIVE AND CONTRACTIVE DISTRACTIONS ON THE WASTING OF TIME ON SUBSEQUENT TASKS¹

DAVID LANDY, KATHLEEN McCUE, AND ELLIOT ARONSON²

University of Texas

It has been demonstrated previously that persons who, because of an "accident," are allowed more time than is necessary to perform an initial task will, of their own accord, spend a greater amount of time on a subsequent similar task than persons who are allowed a minimum amount of time to perform the initial task. The present experiment replicates this excess time effect with the amount of time Ss spend working on the initial task being manipulated by means of a distracting confederate. During performance of the initial task, a confederate imposed on one-half of Ss a contractive distraction (one relatively decreasing work time) and on one-half of Ss a protractive distraction (one relatively increasing work time). In a subsequent work situation, Ss performed a similar task in the absence of the confederate and were allowed to work as long as they chose. The Ss in the Protractive Distraction condition spent significantly more time working on the initial task and significantly more time on the second task than did Ss in the Contractive Distraction condition. This not only demonstrates mediation of the excess time effect by distraction, but also eliminates a possible artifact in the previous experiments—that *E*'s instructions conveyed differential time norms to Ss and were thus responsible for the effect.

In an experimental demonstration of Parkinson's Law that work expands to fill the time available, Aronson and Landy (1967) allowed Ss either 5 or 15 min. to perform a task which could be completed in 5 min. The Ss who were allowed the extra time spent a significantly greater amount of time actually working on the task than those who were allowed minimum time. In addition, Ss were subsequently presented with a similar task and allowed to work at their own pace. Again, Ss who were allowed excess time on the initial task spent significantly more time performing this second task than those who were allowed minimum time on the initial task. This latter phenomenon, termed "the excess time effect" had previously been demonstrated by Aronson and Gerard (1966).

The present experiment is an attempt to eliminate a possible artifact which might have accounted for the excess time effect in the Aronson-Gerard and Aronson-Landy experiments. In both of these experiments the allocation of time was manipulated in the following manner: After *E* had explained the first task to *S*, the departmental secretary burst into the experimental room and urgently reminded *E* that he had promised to help a professor set up some apparatus, and that it would require about 5 min. (in the Minimum Time condition) or 15 min. (in the Excess Time condition). After some hesitation, *E* agreed to help. He suggested that *S* work on the task during *E*'s absence, and as he left the room, *E* said that he would be back in 5 (or 15) min. In describing these studies, the investigators reasoned that, since the time allotted for the tasks was made to appear accidental and arbitrary, it could not have reflected *E*'s judgment of how much time he himself thought the task should consume. However, it is possible that, despite this procedure, differential time norms may have been conveyed by *E* as to what constituted an ap-

¹ This experiment was supported by a grant from the National Institute of Mental Health (MH 12357) to Elliot Aronson. The authors would like to thank Ira Levy, who served as the confederate throughout the experiment.

² Requests for reprints should be sent to Elliot Aronson, Department of Psychology, University of Texas, Austin, Texas 78712.

propriate time to spend on the given task (see Gerard, 1967). Specifically, in the Five-Minute condition in the Aronson-Gerard (1966) and Aronson-Landy (1967) experiments, by stating that he would return in 5 min., *E* may have implied that he expected *S* to be finished within 5 min. In the present experiment, this possibility is eliminated in the following manner: All *Ss* are allowed excess time (15 min.) to perform the initial task; the amount of time which *Ss* actually spend working on the task is manipulated through the actions of a distracting confederate.

Distraction is defined as any alternative activity which competes for a worker's attention during the time which he has at his disposal to complete a specific task. Distraction can be classified into two categories: protractive distractions and contractive distractions. Protractive distractions tend to diffuse or stretch out the amount of time a person will spend working on a given task; they divert attention from the given task without completely halting the work process. Contractive distractions have a relatively opposite effect, tending to decrease or concentrate the amount of time a person will spend working on a task. This latter type of distraction is one which is clearly incompatible with working on the given task, that is, a worker is unable to perform his assigned task and simultaneously attend to the distraction. Therefore, he tends to work harder on the assigned task so that he will finish it more quickly, enabling him to engage in or attend to the distracting activity. By behaving in a predetermined manner, a confederate should be able to impose on *Ss* either a protractive or contractive distraction and thus manipulate the amount of time *Ss* will spend actually working in the initial task.

The Aronson-Gerard and Aronson-Landy experiments demonstrate that the amount of excess time an individual spends working on a task in one situation affects the amount of time he will spend working on a similar task on subsequent occasions. Thus, it was predicted that workers exposed to a protractive distraction in an initial work situation will spend more time at a subsequent similar task in which there is no distraction than workers who are exposed to a contractive distraction

in the initial work situation. Even in a situation where excess time to complete a task is provided to two groups of workers, those who are confronted with a protractive distraction will spend more time completing a subsequent similar task than those who are confronted with a contractive distraction during performance of the first task. In effect, then, the present experiment is a conceptual replication of the Aronson-Gerard and Aronson-Landy experiments. It is designed to demonstrate more clearly that the excess time effect is not merely an artifact of norms implied by the experimental instructions, but is a function of the amount of time consumed in the initial performance of the task.

METHOD

The *Ss* were 42 male undergraduates³ who were required to participate in psychological research in order to accumulate credits required in their introductory psychology class. Individual *Ss* were randomly assigned to one of the two experimental conditions, that is, either Protractive Distraction or Contractive Distraction.

The *E* met *S* and the confederate (posing as another *S*) in a departmental office and introduced herself as Dr. Aronson's assistant. She led them both to an experimental cubicle in which there was a desk and three chairs. When *S* and the confederate were both seated, *E* apologized for the crowded conditions and explained that usually each *S* was alone in a separate cubicle, but that all of the other cubicles were in use at that moment. She said that she hoped that another cubicle would be available soon.

At this point the nature of the task was introduced. The *E* informed *S* and the confederate that they were not participating in an actual experiment; rather, they would be helping *E* to prepare some demonstration materials for Dr. Aronson's social psychology class. She explained that these demonstrations involved interpersonal perception. She also told them that they would receive full experimental credit for their participation, even though, strictly speaking, it was not an actual experiment. She then gave both *S* and the confederate three envelopes, each containing a set of eight photographs. The three sets of pictures given to *S* consisted of ordinary photographs of men. The pictures given to the confederate, on the other hand, were all of women, some of whom were partially nude or in provocative poses. The *E* explained that the task was to rank the first set of photographs according to *S's* own perception

³ Two *Ss* in the Protractive Distraction condition were discarded: One because he left the experimental room during the first work period, and the other because he had previously met the confederate and was suspicious about the distraction.

of the intelligence of the people in the photographs, the second set in terms of perceived warmth, and the third set according to perceived honesty.

At this point, in the interests of strengthening credibility, the confederate asked a question of clarification: "You mean all we do is look at these cards and put them in correct order?" The *E* nodded, while assuring *Ss* that there were no 'right' or 'wrong' answers.

The experiment had been planned so that at this time another confederate would knock on the cubicle door. He told *E* that she was needed to help set up some apparatus and that it would "Only take about 15 minutes." After some hesitation, *E* said that she would be right there. She then apologized to *S* and confederate for the interruption. She explained that she had to leave but suggested that they work on the task while she was gone. As she left the room, she said, "I'll be back in about 15 minutes."

The confederate then exposed *S* to one of the two experimental manipulations (distractions). The confederate had randomly assigned *S* to a given condition before he had met *S* and without *E*'s knowledge.

Contractive Distraction condition. In this condition the confederate allowed *S* to work on his task for 1 min. and then commented, "Hey, are all yours girls?" He then showed *S* one of the nudes. All *Ss* expressed a desire to see the rest of the confederate's pictures, but the confederate told *S* that he should hurry up and finish his rankings and then he could look over the nude photographs at his leisure. Twice more during the task the confederate urged *S* to hurry. When *S* had finished ranking all three sets of photographs, the confederate allowed *S* to see the pictures he had been ranking. He then engaged *S* in a conversation to keep him from going back to recheck his own task.

Protractive Distraction condition. In this condition, the confederate allowed *Ss* to work without interruption for about 1 min. He then began to distract *S* by shoving one of his pictures in front of him, giggling, and making the following comments: "How can you rank a nude girl on intelligence?" "Would I ever like to snuggle up with this one!" "What a build!" "Say let me see your pictures," etc. These interruptions were made at intervals of 15-30 sec. In both the Contractive and Protractive Distraction conditions the confederate finished his own rankings just after *S*.

During this procedure *E* had entered an adjacent cubicle. A peep-hole device enabled her to measure the amount of time *Ss* actually spent working on the task and the distribution of time spent on the task over the 15-min. period during which *E* was absent. When 15 min. had elapsed, *E* returned to the experimental cubicle and again apologized for the previous interruption. She then asked *S* and the confederate to read off the numbers on the back of the photographs so that she could record their rankings. This was done simply to uphold the cover story. After recording the rankings, she said that she had managed to find another vacant cubicle so that each person could work in his own room. She handed *S*

another three sets of photographs (these were similar in nature to the first three sets ranked by *S*) and explained that he was to rank the first set on intelligence, the second on warmth, and the third on honesty, as he had done with the previous photographs. She then requested that the confederate come with her, saying that he, too, would work on a similar task in the newly freed cubicle. As an afterthought, she turned back to *S* and said that after she had directed the confederate to the newly freed cubicle, she wanted to finish setting up the apparatus she had been working on. She assured *S* that she would be just down the hall and suggested that he simply knock loudly on the door of his experimental cubicle when he had finished ranking the pictures; she would come down to record his rankings. The *E* and confederate then left the room. The *E* activated a stop-watch and again entered the adjacent cubicle. When *S* knocked on the door, *E* recorded the elapsed time. She then returned to the experimental cubicle and recorded *S*'s rankings. The *S* was then interviewed. After satisfying herself that *S* was not suspicious about the role of the confederate or about the true purpose of the experiment, *E* fully explained the experiment to *S*, the nature of the deception, and the necessity for employing it.

RESULTS AND DISCUSSION

Before looking at the primary data, it would be prudent to examine our check on the manipulations. In the initial work period during which the confederate manipulated the distraction variable, it was assumed that an *S* was actually working on the task if he was observed to be looking at the photographs which he was instructed to rank. In the initial 15-min. work period the 20 *Ss* in the Contractive Distraction condition spent a mean time of 217 sec. (approximately $3\frac{1}{2}$ min.) actually working on the task while the 20 *Ss* in the Protractive Distraction condition spent a mean time of 395 sec. (approximately $6\frac{1}{2}$ min.) actually working on the task. This difference is significant at beyond the .001 level of probability ($F = 42.25$; $df = 1, 39$) and indicates that the manipulation was effective; that is, Protractive *Ss* consumed reliably more time than Contractive *Ss*.

It will be recalled that during the performance of the second task, *Ss* were allowed (a) to work at their own pace; that is, for as much time as they wished to spend on the task, and (b) to work by themselves; that is, in the absence of the confederate. It was predicted that under these circumstances *Ss* who had been exposed to a protractive distraction

during their performance of the initial task would spend more time working on the second task than Ss who had been exposed to a contractive distraction in the initial work situation. The data clearly support this prediction. The Ss in the Protractive Distraction condition spent a mean time of 279 sec. (more than $4\frac{1}{2}$ min.) working on the second task while Ss in the Contractive Distraction condition spent a mean time of 210 sec. ($3\frac{1}{2}$ min.) working on the same second task under identical conditions. An analysis of variance showed this difference to be significant at beyond the .025 level ($F = 5.44$; $df = 1, 39$).⁴

Since all Ss were allowed the same amount of excess time in which to perform the initial task (15 min.), and since *E* did not know to which condition Ss had been assigned until after she left the room, it is impossible for the excess time effect to have been an artifact of demands or norms implied by *E*. Because these data are parallel to the results obtained by Aronson and Gerard (1966) and Aronson and Landy (1967) using a totally different manipulation, our confidence is increased regarding the "realness" and generality of the effect.

The availability of excess time is a condition existing in many work situations. One of

the important determinants of just how much of the available excess time will be utilized to work on the task at hand will be the presence of distractions in the work situation. The present experiment demonstrates that distractions in one work situation, in which a person is provided with excess time to perform a task, will affect the amount of time that he will spend working on a subsequent similar task in which there is no distraction and in which the person works at his own pace. Whether the amount of time spent on the subsequent task tends to increase or decrease will depend on the nature of the distraction to which the worker is exposed in the initial work situation. Protractive distractions in the initial work situation tend to increase the amount of time spent on a similar task in a subsequent situation while contractive distractions tend to decrease it.

REFERENCES

- ARONSON, E., & GERARD, E. Beyond Parkinson's law: The effect of excess time on subsequent performance. *Journal of Personality and Social Psychology*, 1966, 3, 336-339.
- ARONSON, E., & LANDY, D. Further steps beyond Parkinson's law: A replication and extension of the excess time effect. *Journal of Experimental Social Psychology*, 1967, 3, 274-285.
- GERARD, E. O. A study of the Parkinson effect of excess performance time. Unpublished doctoral dissertation, University of Minnesota, June 1967.
- MILLER, N. Liberalization of basic S-R concepts. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959.

(Received May 10, 1968)

⁴ When the analysis was performed on speed scores, that is, on reciprocals of time in seconds multiplied by 100 (a procedure which is recommended by Miller [1959, p. 283] to compensate for the frequently skewed distribution of time measures), an F of 4.15 was obtained ($df = 1, 38$; $p < .05$).

EXAMINATION OF SOME MEASURES OF CREATIVE ABILITY BY THE MULTITRAIT-MULTIMETHOD MATRIX

P. GOODMAN,¹ J. FURCON, AND J. ROSE

*Graduate School of Business
University of Chicago*

Measures of six traits of creative ability were examined by the multitrait-multimethod matrix. Structured tests, interviews, and supervisory ratings were gathered from 63 scientists and engineers in a research laboratory. There was no evidence of convergent and discriminant validity for the measures of creative ability, although two control traits in the matrix—Job Involvement and Time Extension—exhibited substantial validity. Implications of the findings were discussed.

The purpose of this study is to assess the convergent and discriminant validity of six different measures of creative ability by the multitrait-multimethod technique (Campbell & Fiske, 1959). This technique is primarily concerned with the adequacy of tests as measures of a construct. It provides information for three important issues: (a) whether the trait can be observed under more than one experimental condition, (b) whether the trait can be meaningfully differentiated from other traits, and (c) how much of the variation between traits can be attributed to characteristics of the trait versus the measure of these traits.

Three methods were used to measure the multiple traits: a paper and pencil test, a structured interview, and a paired comparison rating procedure.

The creative ability traits were Sensitivity to Problems, Remote Association, Originality, Ideational Fluency, Spontaneous Flexibility, and Semantic Redefinition. They were selected on the basis of their relevance in the scientific research process and the empirical evidence supporting their validity (Guilford, 1959, 1967 [pp. 162-166]; Mednick, 1962). The selection of the structured test for each trait was based on the degree of empirical validity evidenced in an adult working population and its face validity for the population in our study (Guilford, 1967, Ch. 6; Mednick, 1967).

¹ Requests for reprints should be sent to Paul S. Goodman, Graduate School of Business, University of Chicago, 5836 Greenwood Avenue, Chicago, Illinois 60637.

METHOD

Sample

Seventy-eight employees in a government research laboratory voluntarily participated in the study. Complete test, interview, and ratings were available for only 63 individuals and only these were used in the present analysis.

Fifty-six percent of the individuals used in the analysis were involved in basic research, 43% in engineering problems, and 1% in the administrative area. Fifty-four percent of the sample had completed or were in the process of completing work on their PhD degrees, 35% possessed Master's or Bachelor's degrees, and the remaining had at least some high school education. The average age of the group was 38; the range was 24-61 yr.

Instruments

The following is a brief description of the selected traits of creativity and the respective tests: Remote Association (RAT) was measured by the *Remote Associates Test*.² The test is based on Mednick's (1962) conceptualization of creativity as a process of combining elements (preferably diverse) into new and useful combinations. Sensitivity to Problems (SP)—the ability to recognize problems—was measured by the *Seeing Problems Test* (Guilford, 1967, p. 106). Ideational Fluency (IF)—the ability to call up ideas wherein quantity, not quality, is emphasized (Guilford, 1967, pp. 142-143)—was measured by the number of acceptable responses made to the *A. C. Test of Creative Ability*. Spontaneous Flexibility (SF)—the ability to produce divergent responses (Guilford, 1967, p. 143)—was also measured by the *A. C. Test*. The scoring for this dimension focused upon the number of different classes of responses rather than the total number of responses. Originality (O)—the ability to produce remote, uncommon, or clever responses (Guilford, 1967, pp. 153-158)—was operationalized by the rela-

² Scores on all tests were arranged to exhibit positive relationships.

TABLE 1
MULTITRAIT-MULTIMETHOD MATRIX OF SELECTED CREATIVITY MEASURES

| | Method 1 TEST | | | | | | | | Method 2 INTERVIEW | | | | | | | | Method 3 RATING | | | | | | | |
|-------------------------|------------------|------|------|------|------|------|------|------|-----------------------|------|------|------|------|------|------|------|--------------------|------|------|------|------|------|------|------|
| | RAT | PI | IF | O | SF | SRD | JI | TE | RAT | PI | IF | O | SF | SRD | JI | TE | RAT | PI | IF | O | SF | SRD | JI | TE |
| Method 1—TEST | | | | | | | | | | | | | | | | | | | | | | | | |
| Remote Association | (87) | | | | | | | | | | | | | | | | | | | | | | | |
| Problem Identification | 22 | (80) | | | | | | | | | | | | | | | | | | | | | | |
| Ideational Fluency | 04 | 64 | (85) | | | | | | | | | | | | | | | | | | | | | |
| Originality | 03 | 50 | 58 | (85) | | | | | | | | | | | | | | | | | | | | |
| Spontaneous Flexibility | 11 | 45 | 67 | 49 | (80) | | | | | | | | | | | | | | | | | | | |
| Semantic Redefinition | 43 | 41 | 33 | 32 | 34 | (84) | | | | | | | | | | | | | | | | | | |
| Job Involvement | -27 | -03 | -06 | -02 | 01 | -14 | (83) | | | | | | | | | | | | | | | | | |
| Time Extension | -12 | 06 | -04 | -01 | 17 | 11 | 11 | (88) | | | | | | | | | | | | | | | | |
| Method 2—INTERVIEW | | | | | | | | | | | | | | | | | | | | | | | | |
| Remote Association | -10 | 14 | 10 | 04 | 06 | -07 | 06 | 06 | (66) | | | | | | | | | | | | | | | |
| Problem Identification | -21 | -12 | -02 | 02 | 10 | -18 | 07 | 13 | -03 | (62) | | | | | | | | | | | | | | |
| Ideational Fluency | -11 | 14 | 19 | 14 | 10 | 01 | 20 | 18 | 06 | 23 | (66) | | | | | | | | | | | | | |
| Originality | -26 | -07 | 16 | 08 | 10 | -15 | 30 | 08 | 08 | 05 | 34 | (76) | | | | | | | | | | | | |
| Spontaneous Flexibility | -16 | 13 | 05 | 12 | 10 | -03 | 15 | 09 | 07 | 39 | 51 | 23 | (85) | | | | | | | | | | | |
| Semantic Redefinition | 05 | 09 | 04 | 03 | 06 | 01 | 14 | 07 | 27 | 13 | -03 | 41 | 19 | (76) | | | | | | | | | | |
| Job Involvement | -35 | -11 | -06 | 03 | 10 | -26 | 61 | 13 | 01 | 23 | 08 | 12 | 21 | 05 | (61) | | | | | | | | | |
| Time Extension | -11 | -01 | 04 | 12 | 15 | 02 | 06 | 69 | 18 | 11 | 18 | 18 | 06 | 02 | 16 | (67) | | | | | | | | |
| Method 3—RATING | | | | | | | | | | | | | | | | | | | | | | | | |
| Remote Association | 27 | 14 | 18 | 16 | 31 | 27 | -10 | 11 | 05 | 02 | 10 | 09 | 12 | 06 | -14 | 11 | (94) | | | | | | | |
| Problem Identification | 13 | 21 | 21 | 10 | 23 | 16 | 04 | 12 | -01 | -02 | 20 | 21 | 09 | 13 | -03 | 05 | 66 | (97) | | | | | | |
| Ideational Fluency | 19 | 05 | 08 | 10 | 17 | 21 | -05 | 03 | 14 | -07 | 19 | 13 | 19 | 04 | -18 | 12 | 85 | 65 | (96) | | | | | |
| Originality | 19 | 10 | 14 | 12 | 03 | 14 | -21 | -02 | 07 | 04 | 25 | 16 | 19 | 04 | -19 | 11 | 71 | 59 | 78 | (94) | | | | |
| Spontaneous Flexibility | 26 | 04 | 09 | 07 | 11 | 09 | -19 | 02 | 10 | 01 | 23 | 05 | 16 | 09 | -19 | 10 | 76 | 52 | 80 | 81 | (98) | | | |
| Semantic Redefinition | 06 | 12 | 14 | 04 | 13 | 04 | -13 | 05 | 19 | -05 | 19 | 17 | 25 | -01 | -12 | 15 | 64 | 49 | 74 | 76 | 72 | (95) | | |
| Job Involvement | 09 | -03 | 07 | -04 | 10 | 03 | 01 | 10 | -07 | 07 | -05 | -17 | -07 | -16 | -05 | 16 | 61 | 63 | 52 | 51 | 46 | 43 | (96) | |
| Time Extension | 10 | 10 | 14 | 03 | 19 | 22 | -21 | 18 | 07 | 00 | -10 | -03 | -05 | 16 | -12 | 12 | 64 | 55 | 56 | 52 | 57 | 48 | 54 | (97) |

tive infrequency of a response to the *A. C. Test*. The last dimension, Semantic Redefinition (SR)—the ability to shift the function of an object or particular object and use it in a new way (Guilford, 1967, p. 181)—was measured by the *Object Synthesis Test*.

Two control traits—Job Involvement and Time Extension—were introduced because they were considered independent of the creative ability traits, a requirement in the multitrait-multimethod matrix. Job Involvement (JI) refers to the degree to which an individual is psychologically identified with his work. Time Extension (TE) refers to the length of future time span which is conceptualized. The test measure of Job Involvement was developed by Lodahl and Kejner (1965), Time Extension by Goodman (1966). Both measures are Likert-type scales.

In the interview the respondent was presented with a series of statements describing the traits under examination and then was asked to describe himself in terms of these statements along a 9-pt. scale. The rating procedure used the same stems as the interview. The rater's task was to compare pairs of individuals in terms of the dimension under consideration.

Reliability estimates for the tests were determined by the split-half procedure adjusted by the Spearman-Brown formula. Since it was not possible to reinterview the same population, a retest coefficient was determined from ■ heterogeneous population: three student groups ($N = 6$; $N = 11$; $N = 9$), ■ group of industrial chemists ($N = 17$), and a group of social scientists employed at a research center

($N = 9$). The median rank order retest coefficient is probably understated because the range of responses was more constricted in this population and it was administered on a group rather than individual basis. The reliability estimate for the paired comparison is an average consistency score for all raters for a given dimension.

Administration Procedures

The tests were administered in a group session which lasted approximately 3 hr. The interview time was approximately 1½ hr. and followed from 2–3 wk. after the test administration. In the rating, supervisors were given a deck of cards containing all possible pairs of individuals they agreed to rate for each trait. The approximate rating time, 45 min., followed from 4 to 6 wk. after the test administration.

RESULTS

Data for the multitrait-multimethod matrix is presented in Table 1. Convergent validity is indicated in the heteromethod-heterotrait blocks by the values in the diagonal (e.g., RAT_1 RAT_2); discriminant validity is indicated to the degree a diagonal value exceeds its corresponding row and column values (e.g., RAT_1 JI_1). Since the matrix requires comparison between independent traits, this

analysis focuses on comparisons between each creative ability trait and the control traits, Job Involvement and Time Extension.

No convergent validity for Remote Association appears in the test-interview and rating-interview blocks. The .27 correlation in the test-rating block does not provide significant evidence for convergent validity because of the halo effect operating in the rating measure.³ Examination of the other creativity measures indicates no evidence of convergent and discriminant validation. Only the control measures—Job Involvement and Time Extension—exhibit substantial convergent and discriminant validity.

Method variance can be illustrated by examining the test and rating heterotrait-monomethod (solid) triangles. For example, consider the degree of association between Problem Identification and Ideational Fluency. The correlation for the same measure of these different traits is greater than the correlation between two different measures of the same trait. These findings indicate that the nature of the measurement process must be contributing to the relationship between the two traits.

Two types of method variance seem to appear. First, since the tests for Problem Identification and Ideational Fluency require listing responses, the nature of the instrument probably accounts for some of the associations between both traits. Second, the high association among all the different traits in the rating triangle indicates a halo effect.

DISCUSSION

Lack of substantial convergent and discriminant validity for the selected measures of creative abilities raises serious questions about the nature of the tests. Before any implications are drawn, alternative explanations for the results in the matrix should be considered.

First, an incorrect matrix design would mitigate any useful interpretations of the results. However, the above matrix does follow the Campbell and Fiske (1959) specifications of independent methods and traits (as indi-

cated by the low values in the test and interview heterotrait-monomethod triangle). Also, the convergent and discriminant validities exhibited in the control trait measures seem to indicate the matrix design is adequate.

Second, the low values may be attributed to distribution or instrument factors. Each distribution was examined for skewness and non-linearity. In cases where another statistical technique was more appropriate, a comparison was made with the values presented in the above matrix; in general there was little difference in magnitude or relative relationships. The reliability of the instruments seems adequate and not a major contributor to the low matrix values. Also, the low intercorrelations in the solid triangles for tests and interviews suggest that the problem of social desirability is not a contributor to the observed results. The halo effect in the ratings, which cancelled the usefulness of the rating-test interviewer comparison, seems more a function of lack of knowledge about the ratee⁴ than some inherent characteristic of the rating technique or some characteristic of its administration (e.g., a relationship with the length of rating time and observed halo).

It may be concluded, then, that the low convergent and discriminant validity of the selected tests of creative ability is not a function of inappropriate application of statistical techniques or of instrument error, but of the operational differences in the tests. This lack of convergence between different measures of the same trait raises some question as to the operational meaning of the tests. The implications of these findings for future research on creativity may be summarized as follows: It will be necessary to extend the use of the multitrait-multimethod matrix to different tests for the traits under consideration, to different methods, and in different populations, before other types of validation work are undertaken.

⁴ Raters participated in the interviews and data from the interviews suggests the raters understood the traits in terms of their own behavior and discriminated between the creative traits and control traits.

REFERENCES

- CAMPBELL, D. T., & FISKE, W. D. Convergent and discriminant validation by the multitrait-multi-

³ The "halo" is indicated by the high correlations among ratings of hypothesized independent traits (i.e., RAT, Job Involvement, and Time Extension).

- method matrix. *Psychological Bulletin*, 1959, **56**, 81-105.
- GOODMAN, P. A study of time perspective: Measurement and correlates. Unpublished doctoral dissertation, Cornell University, 1966.
- GUILFORD, J. P. *The nature of human intelligence*. New York: McGraw-Hill, 1967.
- GUILFORD, J. P. Traits of creativity. In Anderson (Ed.), *Creativity and its cultivation*. New York: Harper, 1959.
- LODAHL, T. M., & KEJNER, M. The definition and measurement of job involvement. *Journal of Applied Psychology*, 1965, **49**, 24-33.
- MEDNICK, S. A. The associative basis of the creative process. *Psychological Review*, 1962, **69**, 220-232.
- MEDNICK, S. A., & MEDNICK, M. T. *Examiner's manual: Remote Associates Test*. Boston: Houghton Mifflin, 1967.

(Received May 13, 1968)

ASYMMETRICAL TRANSFER IN READING TEXTS PRODUCED BY TELEPRINTER AND BY TYPEWRITER¹

E. C. POULTON²

Applied Psychology Research Unit, Cambridge, England

Sixty good readers and 66 poorer readers were given 90 sec. to read passages of about 450 words. They had then to answer 10 open-ended questions on the content. The passages and questions were reproduced in Siemens teleprinter capitals with triple spacing, and in IBM 72 elite upper and lowercase letters with single spacing. Length of lines averaged 6.5 in. After reading two passages in one style, two-thirds of the readers were transferred to a passage in the other style. The remaining one-third read one style throughout. The poorer readers found elite easier to comprehend than Siemens ($p < .05$). They showed positive transfer when reading Siemens after elite ($p < .05$), and negative transfer when reading elite after Siemens ($p < .05$). The good readers did not find elite any easier than Siemens, and showed no reliable transfer effects. The product moment correlation between the scores on the first two passages was only .32, because the rate of comprehension depends upon previous knowledge. After correcting for attenuation, the correlation between the pooled rate of comprehension of the first two passages and Part 2 of the Tinker speed of reading test was .84.

Offices may handle both teleprinted and typewritten material. At present teleprinters still print only in capitals. The letters of the Siemens teleprinter have a height of 8 points (2.7 mm.), and are typed in succession 10 letters to the inch. The minimal vertical spacing between lines (which corresponds to leading) is 4 points (1.3 mm.). Sometimes double or even treble spacing is used between lines. This results in only three or two lines of teletype per inch of paper.

In contrast, most typewritten material is typed in lowercase letters, with capitals only at the start of sentences and for proper names. The letters of elite typewriters have an x-height (the height of the rounded parts of the letters, excluding the ascenders and descend-

ers) of only 5 points (1.7 mm.), although the total letter height is 9 points (3.0 mm.). The letters are typed in succession 12 letters to the inch. The usual single spacing gives a vertical separation between lines (corresponding to leading) of only 3 points (1.0 mm.). This makes 6 lines of typewriting per inch of paper.

There can thus be quite a large difference between teleprinted and typewritten material. In offices which handle quantities of both kinds of material, complaints have been made that single-spaced elite type is too small to read comfortably. These complaints could not have been predicted from the results of previous experiments. Tinker and Paterson (1928) and Tinker (1955) found that normally printed lowercase texts were easier to read than texts printed all in capitals. Poulton and Brown (1968, Table 2) found that texts typewritten in elite type were easier to read than texts teleprinted in Siemens capitals, or typed in elite or pica capitals. The experimental findings that lowercase texts are the easier to read conflicts with the complaints that elite type is too small. It suggests that the complaints may be based upon transfer effects. People who have been reading for most of the day material teleprinted in large capital letters with wide spacing between lines may

¹ The problem of transfer was raised by M. A. G. Howgate of the British Government Communications Headquarters, who supplied the Siemens teleprinted material used in the experiment. The passages were written and pretested by C. H. Brock. Experimental Ss of the requisite reading abilities were provided by A. J. Hull. P. M. E. Altham advised on the design of the experiment and on the analysis of the results. K. Tayler ran the correlations. Financial support from the British Medical Research Council is also gratefully acknowledged.

² Requests for reprints should be sent to the author, Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

show negative transfer when given material to read typed in single-spaced elite letters.

There is some evidence which suggests that this could happen. Fox (1963) compared material typed in standard elite with the same material typed in the small-sized capitals of Gothic elite. When read first, the material typed in standard elite was read 18% more quickly than the material typed in Gothic. But when read second, after reading material typed in the other style, the material in the small-sized Gothic capitals was read 20% more quickly than the material in the standard elite. If not due to chance differences between the two groups of readers, this is a two-way asymmetrical transfer effect (Poulton & Freeman, 1966). Texts in small capitals were read more quickly after reading texts in standard elite letters. This is the usual positive transfer or practice effect. Texts in standard elite were read more slowly after reading texts in Gothic capitals. This is an unexpected negative transfer effect, which corresponds in direction to the complaints just described.

The present experiment was designed to investigate this problem. Control groups stuck throughout to a single condition, either Siemens teleprinted material or material typed in elite (see Table 1). Experimental groups practiced on one kind of material, and were then transferred to the other kind. To determine whether negative transfer was due simply to lack of familiarity with material in a particular style, half of the readers in each experimental group were given to read first a passage in the style they were eventually to be transferred to. The other half of the readers met the style for the first time during the experiment in the transfer condition.

METHOD

Materials

Three passages, each of about 450 words, were written on aspects of Roman life: farming, soldiering, and social activity. Sentences were kept short and to the point. Each passage was reproduced twice in lines which averaged 6.5 in. in length. It was teleprinted with triple spacing, using the Siemens all-capitals typeface. And it was typed single-spaced using an IBM 72 elite electric typewriter. Both reproductions were on 4-ply teleprinter paper rolls which have carbon paper interleaved. The top copies,

which were on tracing paper, were used to make the dyeline copies for the experiment.

There were 10 open-ended questions on each passage to test for comprehension. A question required only a few words to answer. The 10 questions were spread evenly over the text, so that a person who had read 80% of the passage would be able to attempt 8 of the 10 questions. After each question there was a line of dots (again triple- or single-spaced) on which the reader had to write his answer. The question sheets were reproduced exactly like the passages: teleprinted with triple spacing and typed in single-spaced elite on 4-ply teleprinter paper. The top copies on tracing paper were used for photocopying by the offset process.

Experimental Design and Subjects

The experimental design is shown in Table 1. It can be regarded as an experiment on six groups, each of 10 good readers, which has been repeated on six groups each of 11 poor readers. A group of good readers was paired with a group of poor readers. Each such pair was treated differently. Previews 1 and 2 involved reading passages without being tested for comprehension.

The two *elite control* groups had elite passages throughout. There were two pairs of *elite transfer* groups. The *prewarned* groups had an elite passage in Preview 1, while the *unwarned* groups did not. Subsequently all four groups had Siemens passages until the final Test 3, when they were given an elite passage to read. The *Siemens control* groups and the *prewarned* and *unwarned Siemens transfer* groups had the corresponding conditions. Everyone read the three test passages in the same order. Thus the relative difficulty of the three passages is confounded with practice effects.

There were 60 good readers who had all scored 50% or more in a previous experiment of a similar nature. There were also 66 poor readers who had all failed to score as much as 50% previously. Ten good readers and 11 poor readers were allocated to each of the six experimental conditions. Allocation was in order of arrival, except for the restriction that each group of good readers had the same number of very good and good readers. Each group of poor readers had the same number of poor and very poor readers.

All 126 readers were members of a panel maintained at the Applied Psychology Research Unit in Cambridge. Just under one-fourth were men. Their ages ranged from 24 to 73 yr. About half wore reading glasses for the experiment. They were paid 7s. 6d. per hr. (about \$.90) for their services plus traveling expenses.

Procedure

Readers allocated to different conditions were tested simultaneously in groups. Those reading elite passages in Tests 1 and 2 were seated separately from those reading Siemens passages. Thus after Preview 1 nobody saw passages in a style other than the style which he was reading until transfer

TABLE 1
EXPERIMENTAL DESIGN AND MEAN PERCENT COMPREHENSION
FOR GOOD AND POOR READERS

| All readers | | | Groups of 10 good readers | | | Groups of 11 poor readers | | |
|-------------------------|-----------|-----------|---------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------------|
| Experimental condition | Preview 1 | Preview 2 | Test 1 | Test 2 | Test 3 | Test 1 | Test 2 | Test 3 |
| Elite | | | | | | | | |
| Control | Elite | Elite | Elite 52 ^a | Elite 65 ^a | Elite 46 ^a | Elite 47 | Elite 47 | Elite 50 ^{b,c} |
| Transfer prewarned | Elite | Siemens | Siemens 51 | Siemens 61 | Elite 50 | Siemens 40 | Siemens 48 | Elite 33 ^b |
| Transfer unwarned | Siemens | Siemens | Siemens 62 | Siemens 64 | Elite 47 | Siemens 37 | Siemens 45 | Elite 34 ^b |
| Siemens | | | | | | | | |
| Control | Siemens | Siemens | Siemens 61 | Siemens 69 ^d | Siemens 50 ^d | Siemens 41 ^a | Siemens 41 ^a | Siemens 20 ^{a,c,e,f} |
| Transfer prewarned | Siemens | Elite | Elite 59 | Elite 68 | Siemens 58 | Elite 45 | Elite 53 | Siemens 35 ^e |
| Transfer unwarned | Elite | Elite | Elite 55 | Elite 67 | Siemens 56 | Elite 45 | Elite 55 | Siemens 30 ^f |
| All conditions combined | | | Elite 55 | Elite 67 | Elite 48 | Elite 45 ^g | Elite 51 ^g | Elite 39 |
| | | | Siemens 58 | Siemens 65 | Siemens 55 | Siemens 40 ^g | Siemens 45 ^g | Siemens 29 |

^a Test 3—Tests 1 and 2 combined $p < .05$ or better.
^b Elite control—Elite transfer, prewarned and unwarned, $p < .05$.
^c Elite control—Siemens control $p < .001$.
^d Test 3—Test 2 $p < .01$.
^e Siemens control—Siemens transfer prewarned $p < .05$.
^f Siemens control—Siemens transfer unwarned $p < .05$ (one-tailed test).
^g Elite—Siemens $p < .05$ on Tests 1 and 2 combined.

in Test 3. The experiment was introduced as a comparison of teleprinted and typewritten texts. Beyond this the readers were told nothing of the aims of the experiment.

In the two preliminary preview conditions passages of about 500 words were read for 2 min. The readers were told that this was merely to acquaint them with the style of print; they would not be questioned on the content. In each test a test passage of about 450 words was studied for 90 sec. Four minutes were allowed subsequently for answering the 10 questions, but most readers did not require as long as this. The experiment lasted about 30 min.

After the experiment everyone was given 10 min. on Part II of the Tinker Speed of Reading Test (Tinker, 1955).

RESULTS

The results are given in Table 1. The groups of 10 good readers are in the middle. The groups of 11 poor readers are on the right. The most difficult of the three passages was chosen for the transfer test, Test 3. This is indicated by the results of three out of the four control groups. The top row of the table

gives the mean comprehension scores of the two groups which read elite type throughout. The 10 good readers in the middle of the table did reliably worse on Test 3 than their average on Tests 1 and 2 ($p < .05$ on a two-tailed Wilcoxon test, Siegel, 1956, p. 75). The 11 poor readers on the right of the table did no worse on Test 3 than on Tests 1 and 2.

The fourth row of the table gives the corresponding data for the two groups which read Siemens throughout. Here it was the poor readers on the right of the table who did reliably worse on Test 3 ($p < .01$). For the good readers in the middle of the table, Test 3 was reliably worse than Test 2 ($p < .01$), but it was not reliably worse than the average of Tests 1 and 2 ($p > .05$).

The good readers showed no reliable difference between elite and Siemens, and no reliable transfer effects on Test 3. The key results concern the poor readers. The data from the poor readers were subjected to analysis of

variance. After removing the differences between individuals and the differences among the three tests ($p < .001$), the residual in the 3×6 matrix given on the right of Table 1 was found to be reliable at the .02 level. Thus there must be reliable differences in the table related to the Siemens and elite lettering, or to the orders of reading the Siemens and elite, or to both the lettering and the orders.

The bottom two rows on the right of the table show that the poor readers scored higher on the elite passages than on the Siemens. For Tests 1 and 2 combined, the difference in favor of elite was reliable at the .05 level on a two-tailed Mann-Whitney U test (Siegel, 1956, p. 116). This is a simple comparison between groups, and is not contaminated by transfer effects.

The right-hand side of the top row of the table gives the mean scores for comprehension of the control group of 11 poor readers who read exclusively elite texts. They did as well on the more difficult third passage as they did on the two previous passages. The right sides of the next two rows give the mean scores of the two groups of poor readers which transferred from Siemens to elite on Test 3. Both groups did rather less well than the control group on Test 1, but even the combined difference was not reliable statistically ($.05 < p < .1$). Taken together the two groups did reliably better on Test 2 ($p < .05$). Their mean scores for comprehension were then about the same as the mean of the control group. On Test 3 they transferred from Siemens to elite. Both groups then had means which were reliably ($p < .05$) less than the mean of the control group.

The right side of the fourth row in the table gives the mean scores for comprehension of the control group of poor readers who read exclusively Siemens texts. They did reliably worse on the difficult third passage than on either of the two previous passages ($p < .01$). The right sides of Rows 5 and 6 give the means of the two groups of poor readers which transferred from elite to Siemens on Test 3. They did a little better than the control group on Test 1. On Test 2 the two groups taken together did reliably better than on Test 1 ($p < .02$). Their means for comprehension were then reliably above the means

of the control group ($p < .05$). On Test 3 they transferred from elite to Siemens. Both groups did reliably better than the control group which had read Siemens throughout ($p < .05$ on two-tailed and one-tailed tests, respectively).

For the good and poor readers combined there was a product moment correlation of .56 between the pooled results on Tests 1 and 2, and Part II of the Tinker Speed of Reading Test. The correlation between the scores on Tests 1 and 2 was only .32. The median reliability of the Tinker test is .86 (Tinker, 1955). After correcting for attenuation by the Spearman-Brown prophecy formula (Guilford, 1936, p. 368), the correlation between rate of comprehension as measured here and Tinker's speed of reading measure was .84.

DISCUSSION

Asymmetrical Transfer

Table 1 shows that the poor readers read Test 3 in elite type reliably *less* effectively after reading Tests 1 and 2 in Siemens than they did after reading Tests 1 and 2 in elite. Whereas they read Test 3 in Siemens reliably *more* effectively after reading Tests 1 and 2 in elite than they did after reading Tests 1 and 2 in Siemens. This is a two-way asymmetrical transfer effect (Poulton & Freeman, 1966). It was shown only by the poor readers, who were reliably worse on Siemens than on elite texts. The good readers were no worse on Siemens than on elite. They did not show reliable transfer effects.

The transfer effect was not due to lack of familiarity with the lettering of Test 3. The two prewarned groups, who had read a practice passage in the same lettering during Preview 1 (see left of Table 1), showed a two-way asymmetrical transfer effect which was at least as large as that shown by the two unwarned groups.

A detailed analysis of the last column of Table 1 suggests that the asymmetrical transfer was due partly to the behavior of the control groups of poor readers. The elite control group did as well on the difficult third test passage as on the two previous passages and had the best average score for compre-

hension, whereas the Siemens control group scored only half as much on the difficult third test passage as on the two previous passages and had the worst score for comprehension. Apparently the combination of the difficult test passage with the difficult Siemens capitals was too much for the poor readers. The large and highly significant difference between the means of the two control groups was a predisposing factor in the two-way asymmetrical transfer effect. Any mean falling in the gap between the two means necessarily showed either positive or negative transfer.

When a balanced factorial design is used, the asymmetrical transfer will reduce the advantage of elite over Siemens, an effect to be expected with poor readers. It could account for the smaller differences between all capitals and normal upper and lowercase texts found by Poulton and Brown (1968) in their Latin-square experiment, compared with their experiment with separate groups.

Asymmetrical transfer between capitals and lowercase may also have occurred among the poorer readers in Tinker and Paterson's (1928) experiment. In this case the difference of about 12% found in favor of the lowercase may underestimate the actual difference in the difficulty of reading. Unfortunately the authors give only the means for good and poor readers combined. The means do not reveal any overall asymmetrical transfer. This may be because the effect on the poorer readers has been masked by the results of the good readers who showed no effects. Positive transfer between the lowercase and all capitals text may have occurred among the poorer readers in Tinker's (1955) repeat experiment—in which case the 12% difference found in favor of lowercase may again have been an underestimation.

Variability in the Rate of Comprehension

The correlation between the rate of comprehension on the first two passages was only .32. This may have resulted partly from individual differences in the way people reacted to the first passage. Everyone had done similar tests before, but not for a year or more. However the principal cause for the low correlation was probably differences in familiarity with the material read. People memorize most

easily what they know already (Poulton, 1957). One passage or part of a passage may contain familiar ideas. It is read quickly and gives a high score for comprehension. Another passage or part of a passage may be quite unfamiliar. It has to be read slowly, or perhaps twice, and even then gives a low score for comprehension. Thus measures of the rate of comprehension necessarily contain a lot of within-individual variability.

The Tinker Speed of Reading Test suffers less from variability in the rate of comprehension, because it is more a test of speed than of comprehension. The test requires the reader to cross out 1 word in about 30. The rate of crossing out words with a pencil must be reflected in the overall speed score. The incongruous word towards the end of each 30-word item can often be spotted with only quite a vague idea of what the item is about. Deep comprehension is not required. The emphasis is on speed, both in crossing out words and in reading rapidly. This may account for the relatively high median reliability of the Tinker test, .86.

However a reading test whose reliability is based upon the speed of locating and crossing out 1 word in 30 is not necessarily representative of normal reading. It is probably closer to skimming or scanning (Poulton, 1967). It seems likely that print which is easy to comprehend is also easy to skim or scan. After correcting for attenuation, the correlation between rate of comprehension as measured here and by the Tinker test was .84. This suggests that the abilities measured by the two tests may be similar. But more evidence is required. The correlation was only .56 before it was corrected for attenuation.

REFERENCES

- FOX, J. G. A comparison of Gothic Elite and Standard Elite typefaces. *Ergonomics*, 1963, 6, 193–198.
- GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936.
- POULTON, E. C. Previous knowledge and memory. *British Journal of Psychology*, 1957, 48, 259–270.
- POULTON, E. C. Skimming (scanning) news items printed in 8-point and 9-point letters. *Ergonomics*, 1967, 10, 713–716.
- POULTON, E. C., & BROWN, C. H. Rate of comprehension of an existing teleprinter output and of possible alternatives. *Journal of Applied Psychology*, 1968, 52, 16–21.

POULTON, E. C., ■ FREEMAN, P. R. Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin*, 1966, 66, 1-8.

SIEGEL, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

TINKER, M. A. Prolonged reading tasks in visual

research. *Journal of Applied Psychology*, 1955, 39, 444-446.

TINKER, M. A., & PATERSON, D. G. Influence of type form on speed of reading. *Journal of Applied Psychology*, 1928, 12, 359-368.

(Received May 20, 1968)

PREDICTING VOCATIONAL SUCCESS FOR NEUROPSYCHIATRIC PATIENTS WITH THE EDWARDS PERSONAL PREFERENCE SCHEDULE

ALLEN GOSS¹

University of Saint Thomas and Texas Research Institute of Mental Sciences

This study investigates the possibility of predicting success-fail vocational outcome from a personality inventory. The Edwards Personal Preference Schedule (EPPS; Edwards, 1959) was administered to a group of 58 psychiatric patients at the time of acceptance to a Vocational Rehabilitation Ward. Score areas of the EPPS utilized in the predictive model varied with the diagnostic categories. The results show that model predictions above the base rate of success are possible for all of the diagnostic groups and that predictive accuracy increases when results are specifically related to particular groups.

Numerous investigators have used the Edwards Personal Preference Schedule (EPPS) in examining relations between hypothesized needs and various behavioral criteria. Working in the area of vocational choice, Norrell and Grater (1960) reasoned that if the self-concept was distorted by a lack of self-awareness, vocational choice would tend to be inappropriate. High and low self-awareness groups were significantly different on Succorance and Order needs from the EPPS. Pool (1965), using a Veterans Administration hospital population of patients with whom vocational counseling was regarded as either effective or ineffective, found that the ineffective group had lower need scores for Intraception and Endurance and higher need scores for Succorance and Autonomy. While only part of Pool's results were similar to those of Norrell and Grater, they both gave dependency interpretations, and they both stressed the importance of psychological needs in making vocational choices.

The focus of the present study was to investigate the utility of EPPS need scores in differentiating patients accepted to a vocational rehabilitation ward with respect to their subsequent employment outcome. An-

other interest was to explore the usefulness of a simple model designed to utilize EPPS score information from the analysis, to place patients in success or fail groups.

In line with results from previous investigations it was hypothesized that those patients who gained suitable employment would have significantly higher scores on Affiliation, Intraception, and Nurturance than those who did not gain employment, and a lower Succorance score.

PROCEDURE

Subjects. The Ss were 58 male neuropsychiatric patients accepted for the vocational rehabilitation program at the Veterans Administration Hospital, Houston, Texas. Patients accepted to this ward receive vocational counseling, industrial therapy through supervised work experience in the hospital, and/or educational therapy, as well as job placement assistance.

Tests. All Ss were given the EPPS as part of the psychological testing just prior or subsequent to admission for the rehabilitation program. The EPPS was not used in the vocational counseling of the patient, nor were the scores readily available to the counselors.

Method. The criterion of employment in the study was related to the discharge of the patient. If the patient was discharged from the ward with a suitable job, he was considered a success. Employment was judged as suitable by the ward vocational counselor if the patient had sufficient skills to meet the demands of the job and if, while on the ward, the patient worked consistently at the job during the several week time-period necessary to accumulate sufficient funds to be self-supporting. If transferred or discharged from the ward in any other category, the patient was considered a failure with respect to the employment criterion of the ward. Each of the

¹Information for this study was gathered while the author was a predoctoral trainee and a postdoctoral fellow at the Veterans Administration Hospital in Houston, Texas. Requests for reprints should be sent to the author, Baylor University, College of Medicine, 1200 Moursund Avenue, Houston, Texas 77025.

15 EPPS need scores was analyzed with respect to the success-failure criterion for the entire sample and for the following diagnostic subgroups: (a) alcoholic (ALC); (b) anxiety-depression (A-D); (c) patients with physical disabilities (P-D); and (d) schizophrenics (SCH). Results from previous investigations with the vocational rehabilitation population (Goss, 1966, 1968; Goss & Pate, 1966, 1967) indicated the need for separate analysis for each of the diagnostic groups. The success-failure diagnostic groups had minimal differences with respect to age, education, and number of previous hospitalizations.

Since the second concern of the study was to explore the possibility of placing patients in success or fail groups which exceed the population base rates through the use of a predictive model utilizing score information, its operation will be briefly described. Items with large P values were included in order to reduce Type 2 (beta) error and to increase power. Scores with P values of less than .20 were included as predictors; however, weights assigned to these scores depended on the P value of the items. Items with a P value between 0 and .10 were weighted +2, 0, or -2, depending on the relation of the item scores to the mean of the success or fail group. If a value was equal to or exceeded the success mean, it was given the positive weight assigned that item; if the value was between the success and fail mean, it was given a zero; and if the value was equal to or below the mean fail value, it was given a minus weight. Similarly, weights of 1 were assigned to items with P values between .10 and .20. By algebraically summing the positive and negative weights of the predictor values, total weight scores of a plus, zero, or minus value were derived for each individual in the various diagnostic groups, and from these values the model predictions were made. Predictively, positive values indicated success, negative values indicated failure, and zero values registered an area of unpredictability.

It was hypothesized that the successful group would have higher scores on Affiliation, Intraception, and Nurturance and a lower score on Succorance than the failures. The results for the total sample indicate that successes generally had a higher score on Succorance (S [success]: 64.58, F [failure]: 47.44; $F = 5.93$, $p < .02$), and a lower score on Deference (S : 44.58, F : 60.56; $F = 5.08$, $p < .03$). Thus, the Succorance score was significant ($p < .02$) in the wrong direction. Affiliation ($p < .77$), Intraception ($p < .53$), and Nurturance ($p < .29$) did not differentiate the population with respect to the employment criterion. Thus EPPS score areas of significance in previous vocational studies employing self-awareness and counseling-suitability criteria were not validated with the employment outcome criterion utilized in this investigation.

Areas which meet the model requirements include the four need scores which differentiated the ALC group: Deference (S : 44.40, F : 71.25; $F = 4.55$, $p < .05$), Succorance (S : 62.30, F : 34.50; $F = 5.00$, $p < .04$), Abasement (S : 52.50, F : 30.37; $F = 1.83$, $p < .20$), and Nurturance (S : 56.50, F : 35.25; $F =$

TABLE 1

MODEL PREDICTION OF EMPLOYMENT SUCCESS FOR
PSYCHIATRIC PATIENTS FROM EPPS SCORES

| Subgroups | No. of patients | Percent- age of patients in model pre- dicted for | Percent- age correct | Ward employ- ment base rate |
|-----------|-----------------|---|----------------------------|---|
| ALC | 18 | 94 | 82 | 48 |
| A-D | 20 | 80 | 81 | 73 |
| P-D | 9 | 67 | 83 | 52 |
| SCH | 11 | 100 | 82 | 54 |
| Total | 58 | 67 | 74 | 58 |

2.09, $p < .17$). The A-D group contributed three need scores: Affiliation (S : 46.77, F : 27.57; $F = 1.79$, $p < .20$), Succorance (S : 60.54, F : 36.14; $F = 2.90$, $p < .11$), and Heterosexuality (S : 72.77, F : 46.43; $F = 6.00$, $p < .02$). The P-D group showed the least differentiation with one score providing minimal contribution, Nurturance (S : 65.50, F : 43.40; $F = 2.33$, $p < .17$). The SCH population showed slight differences on Affiliation (S : 39.83, F : 67.00; $F = 2.16$, $p < .18$), and on Aggression (S : 59.00, F : 18.80; $F = 6.82$, $p < .03$).

An analysis of variance between the scores of the four diagnostic groups indicated that there were two areas which significantly differentiated these groups: Achievement ($p < .04$) and Order ($p < .01$). These differences were due to the low A-D group mean scores, indicating that the lowest need Achievement and Order scores occurred in the group with the highest percentages of success.

The model results are presented in Table 1. These results show that predictions above the base rates of success are possible for all populations and that the accuracy and percentage of these predictions increase when the results are specific to particular populations. From these results the authors conclude that attempts to predict vocation outcomes—or any other performance-specific behavior—is a reasonable task provided that the authors wish to make predictions for fairly homogeneous groups and that they can gain behavior-specific criterion information. The present results were related to the employment outcome of patients; they compare very favorably with the results of previous investigations. The relation between constructs such as dependency, self-awareness, and effective vocational counseling are clearly less than isomorphic to the objective requirements of a workaday world.

REFERENCES

- EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule*. (Rev. ed.) New York: Psychological Corporation, 1959.

- Goss, A. *Predicting work success for patients on an industrial rehabilitation ward in a neuropsychiatric setting*. (Doctoral dissertation, University of Texas) Ann Arbor, Mich.: University Microfilms, 1966. No. 66-14381.
- Goss, A. Importance of diagnostic categories in evaluating psychological data. *Journal of Counseling Psychology*, 1968, 15, 476-478.
- Goss, A., & PATE, K. Work attitudes and industrial adjustment of psychiatric patients. *Newsletter for Research in Psychology*, 1966, 8, 9-10.
- Goss, A., & PATE, K. Work attitudes and industrial adjustment of psychiatric patients (Part II). *Newsletter for Research in Psychology*, 1967, 9, 8-9.
- NORRELL, G., & GRATER, H. Interest awareness as an aspect of self-awareness. *Journal of Counseling Psychology*, 1960, 7, 289-292.
- POOL, D. A. The relation of personality needs to vocational counseling outcome. *Journal of Counseling Psychology*, 1965, 12, 23-27.

(Received May 21, 1968)

RELATION BETWEEN EMBEDDED FIGURES TEST PERFORMANCE AND SIMULATOR BEHAVIOR

GERALD V. BARRETT

Management Research Center, University of Rochester

AND CARL L. THORNTON AND PATRICK A. CABE

Goodyear Aerospace Corporation, Akron, Ohio

Relationships previously found with reaction to an emergency situation and simulator sickness compared to the Rod and Frame Test (RFT) measure of perceptual style were extended using a second perceptual style measure, an embedded figures test (EFT). The RFT was significantly related to emergency behavior and to simulator sickness. The EFT was significantly related to emergency behavior but not to the simulator sickness. Implications for the use of both tests in the prediction of driving behavior are discussed.

Barrett and Thornton (1968) found correlations from .55 to .75 between perceptual style as measured by Series 3 (S_3) of the Rod and Frame Test (RFT) and emergency behavior in an automobile simulator. Witkin, Lewis, Hertzman, Machover, Meissner, and Wapner (1954) found that there was a relationship between S_3 and the Embedded Figures Test (EFT). It was possible, therefore, that the EFT would also be related to emergency behavior. Logically, the task of visually extracting a geometric pattern from a complex pattern (EFT) was similar to the emergency task of detecting a pedestrian (in a complex background) moving into the path of a vehicle.

Barrett and Thornton (1968) also found that good S_3 performance was related to simulator sickness. Since in the simulator the visual impression of motion on the screen was not accompanied by any physical motion, sensitivity to visual-kinesthetic conflict (RFT) may explain the phenomenon. Also, the ability to disembed (EFT) the conflicting cues may be important.

Two hypotheses, then, were tested (*a*) that there would be a significant relationship between EFT performance and behavior in an emergency situation; (*b*) that there would be a significant relationship between EFT performance and experience of simulator sickness.

¹ Requests for reprints should be sent to Carl L. Thornton, Department 459 Plant H, Life Sciences Research Department, Goodyear Aerospace Corporation, 1210 Massillon Road, Akron, Ohio 44315.

METHOD

Subjects

A random sample of 50 male Ss (aged 30-45) were selected from approximately 1200 employees in a division of an aerospace corporation. Some Ss developed simulator sickness with 26 Ss leaving before the emergency trial. The data for 3 Ss was not used, 2 because of an error in procedure and 1 because he became aware of the purpose of the study. 20 of 21 Ss were given the RFT 6 mo. later. Six months after that, 18 of the 20 Ss remaining were able to be retested on the EFT. For the simulator sickness comparison, 37 of 46 Ss tested on the RFT were available for retesting with the EFT.

Apparatus

A standard EFT (Form Cf-1) supplied by the Educational Testing Service was used (French, Ekstrom, & Price, 1963). The simulator consisted of a stationary automobile with a projection screen in front of the windshield. The visual scene was obtained from a television camera mounted on a movable track over a scale model highway. The camera moved in direct response to the accelerator, brake, and steering movements in the automobile. In this way S had complete control of changing the visual scene.

Two items of a questionnaire, given to Ss 6 mo. after simulator operation, were concerned with subjective estimates of simulator sickness. The Ss were asked to rate the discomfort they experienced while operating the simulator and to estimate how long after simulator operation that discomfort persisted.

Procedure

The EFT was administered in two equal lengths (16 items in two 10-min. periods). Score was the number of items found in 20 min. The S's task in the emergency behavior study was to respond appropriately when a human-like dummy emerged suddenly onto the highway. The Ss were unaware that

the sequence of events was to occur. Sickness measures were obtained by use of a questionnaire approximately 6 mo. after the simulator trials.

RESULTS

The results of the EFT were compared to two emergency criteria, initial brake reaction time and deceleration rate. The $y = x$ reaction time-RFT comparison had yielded a Pearson r of .61, while the $1/y = x$ deceleration rate-RFT comparison yielded an r of .74 (Barrett & Thornton, 1969). As Thornton and Richards (1968) have pointed out, the comparison of time and rate scores necessitates reciprocal data transformations. Since the EFT used in the present study has a rate score, reciprocals were used. The $y = 1/x$ reaction time-EFT correlation was .54 ($p < .03$); the $1/y = 1/x$ deceleration rate-EFT correlation was .49 ($p < .05$). The RFT- $1/y$ -EFT correlation was .83. Although the correlations were lower, the relations were much better than those obtained with Series 1 and 2 of the RFT.

The highest correlation for the RFT-simulator sickness comparison had been .33-.55. For the EFT comparison they dropped to .10-.29. The first hypothesis, then, was confirmed, but not the second.

DISCUSSION

Lower correlations between EFT and criteria scores than were found between S_3 and the same criteria were the result of a number of factors: (a) the time between measurements was 1 yr. for the EFT and 6 mo. for the RFT. In a year's time perceptual style, as measured by the EFT, may have changed. (b) The EFT is not as reliable as the RFT (Witkin et al., 1954). The RFT is an extremely reliable test ($r = .95$) while over even

short periods of time the EFT reliability only is moderate ($r = .6-.9$). Split-half reliability for the present sample was only about .58. (c) EFT and RFT tap slightly different perceptual dimensions. While the correlation between the RFT series and EFT is usually significant, only 36%-74% of the variance is common. It is evident that perceptual style is not a unitary construct. (d) The EFT sample was smaller than the RFT sample. It is conceivable that the dropping of several Ss affected the results.

Despite these drawbacks it is encouraging that a significant relationship was found between emergency behavior and EFT performance. It is suggested that both the RFT and EFT be given immediately after gathering emergency behavior data. Both should be related to quickness of response and may be combined as powerful tools for prediction of driving behavior. Other situations where quick detection and recognition are necessary may also be studied fruitfully with these two instruments.

REFERENCES

- BARRETT, G. V., & THORNTON, C. L. The relationship between perceptual style and emergency behavior. *Journal of Applied Psychology*, 1968, 52, 169-176.
- BARRETT, G. V., & THORNTON, C. L. The relationship between perceptual style and simulator sickness. *Journal of Applied Psychology*, 1969, 52, 304-308.
- FRENCH, J. W., EKSTROM, R. B., & PRICE, L. A. *Manual for kit of reference tests for cognitive factors*. Princeton: Educational Testing Service, 1963.
- THORNTON, C. L., & RICHARDS, B. L. The use of reciprocals in comparing group and individually administered tests. *Educational and Psychological Measurement*, 1969, in press.
- WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

(Received June 21, 1968)

journal of applied psychology

Table of Contents

| | | |
|--|--|-----|
| Analyzing the Expert Judge: A Descriptive Study of a Stockbroker's Decision Processes. | Paul Slovic | 255 |
| Communication Length as an Index of Communicator Attitude: A Replication. | Arthur N. Wiens, Russell H. Jackson, Thomas S. Managh, and Joseph D. Matarazzo | 264 |
| Biographical Correlates of Artistic and Literary Creativity in Adolescent Girls. | Anne Anastasi and Charles E. Schaefer | 267 |
| Perceived Need Satisfactions of Domestic versus Overseas Managers. | John M. Ivancevich | 274 |
| Sources of Variation in Job and Life Satisfaction: The Role of Community and Job-Related Variables. | Charles J. Hulin | 279 |
| Vocational Choice and Professional Experience as Factors in Occupational Image. | Edmond Marks and Sam C. Webb | 292 |
| Color Coding Effects in Compatible and Noncompatible Display-Control Arrangements. | Gary K. Poock | 301 |
| SVIB Interests and Product Preferences. | Robert A. Peterson and Allan L. Pennington | 304 |
| Effectiveness of SVIB Academic Interest Scales in Predicting College Achievement. | Richard W. Johnson | 309 |
| Rated Acceptability of Mineral Taste in Water: II. Combinatorial Effects of Ions on Quality and Action Tendency Ratings. | William H. Bruvold and William R. Gaffey | 317 |
| Personality Inventory Correlates of Creativity among Architects. | Wallace B. Hall and Donald W. MacKinnon | 322 |
| Validity, Predictive Efficiency, and Practical Significance of Selection Tests. | Ervin W. Curtis and Edward F. Alf | 327 |
| Group and Individual Effects in Problem Solving. | George S. Rotter and Stephen M. Portugal | 338 |
| List of Manuscripts Accepted. | | 342 |

KENNETH E. CLARE
University of Rochester

Consulting Editors

GEORGE E. BRIGGS
Ohio State University

MARVIN D. DUNNETTE
University of Minnesota

NORMAN FREDERIKSEN
Educational Testing Service

LEONARD D. GOODSTEIN
University of

EDWIN R. HENRY
University of Rochester

JOHN HOLLAND
American College Testing
Program

CLIFFORD E. JURGENSEN
Minneapolis Gas Company

DONALD G. MARQUIS
Massachusetts Institute of
Technology

LAURENCE S. MCGAUGHAN
University of Houston

QUINN MCNEMAR
University of Texas

WARREN T. NORMAN
University of Michigan

HAROLD F. ROTHE
Beloit Corporation

THOMAS A. RYAN
Cornell University

ALEXANDER G. WESMAN
Psychological Corporation

CLARK L. WILSON
Instituto Centroamericano de
Administracion de Empresas

This journal gives primary consideration to original investigations in any field of applied psychology except clinical psychology, although a descriptive or theoretical article may be accepted if it represents a special contribution in an applied field. Quantitative investigations of interest or value to psychologists working in the following broad fields will be considered: personnel research in business, industry, and government; engineering psychology; industrial working conditions; research on opinion and morale factors; job analysis and classification research; market and advertising research; vocational and educational prognosis, diagnosis, and guidance at the secondary and college level.

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or they may be obtained from the Editor or from the APA Central Office.

Manuscripts should be addressed to the Editor:

Dr. Kenneth E. Clark, Dean
College of Arts and Science
302 Morey Hall
University of Rochester
Rochester, N. Y. 14627

Authors are strongly advised to follow the general directions given in the *Publication Manual of the American Psychological Association* (1967 Revision). Special attention should be given to the section on the preparation of the references (p. 39). In reference lists, give journal titles in full; do not abbreviate. *All copy must be double spaced, including the references.* All manuscripts should be submitted in *duplicate*, one of which should be an original typed copy; author's name should appear only on title page. Dittoed and mimeographed copies are not acceptable and will not be considered. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail and to check carefully the typing of the final copy.

The following policies govern reprinting or other forms of reproduction of APA-copyrighted journal materials. Permission by the American Psychological Association must be requested for reprinting all tables, figures, and text of more than 500 words (total from one article) in length. There will be a charge of \$5.00 per page for such permission except when the requester is the author of the article to be reprinted, or the article is to be reproduced in a limited number of copies solely for instructional purposes.

Published bimonthly by the
American Psychological Association
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W.
Washington, D. C. 20036

\$10.00 per volume

\$2.00 per issue

HELEN ORR
Managing Editor

CLAIRE PILLIOD
Technical Editor

HENRY S. SALMON
Advertising and
Business Manager

ROBERT D. GORMAN
Subscription Manager

Subscriptions, orders, and business communications should be addressed to the American Psychological Association, 1200 Seventeenth St., N. W., Washington, D. C. 20036. Address changes must reach the subscription office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Second-class postage paid at Lancaster, Pa.

ANALYZING THE EXPERT JUDGE:

A DESCRIPTIVE STUDY OF A STOCKBROKER'S DECISION PROCESSES¹

PAUL SLOVIC²

Oregon Research Institute

This study illustrates an analysis-of-variance technique for describing the use of information by persons making complex judgments. The Ss were two stockbrokers who rated the growth potential of stocks on the basis of 11 factors taken from Standard & Poor's reports. The technique proved capable of providing a precise quantitative description of configural and nonconfigural information utilization. Each broker exhibited a substantial amount of configural processing. The technique appears to have promise for providing experts with insight into their own processes and for teaching and evaluating "student" judges.

The task of the expert judge, no matter what his occupation—military officer, detective, businessman, physician, clinical psychologist, financial analyst, etc.—requires him to combine items of information from a number of different sources into a decision or judgment. The key to the expert's success resides in his ability to interpret and integrate information appropriately. This means he must weigh items of information differentially, according to their relevance, and must be able to qualify his interpretations of a given fact when other considerations make such qualification necessary.

¹ This research was supported by Grants MH 04439 and MH 12972 from the United States Public Health Service. Computing assistance was obtained from the Health Sciences Computing Facility, University of California, Los Angeles. Portions of this work were presented at the meetings of the Western Psychological Association, San Diego, March, 1968.

² The author wishes to thank Terry Ashwill for his invaluable assistance in the design of the study and for his participation as an S, Robert Kraus for serving as the second S, Jerry Solomon and Russel Geiseman for their assistance in analyzing the data, and Sarah Lichtenstein and Leonard Rorer for their comments on the manuscript.

Requests for reprints should be sent to the author, Oregon Research Institute, P. O. Box 3196, Eugene, Oregon 97403.

There is no need to dwell upon the tremendous importance of being able to understand and describe how the expert uses information. However, such understanding does not come easily. All too often expert judgment is regarded as a mysterious, intuitive phenomenon, incapable of being described precisely. For example, Lusted (1960) relates a story about a radiologist famed for his diagnostic ability. Once, when he was questioned as to why he thought a particular shadow on an X-ray was a metastatic lesion, the physician replied, "Because it looks like it!" At the other extreme is the expert who instructs others in the art of emulating his judgments by reeling off the dozens of factors that he takes into consideration, each accompanied by an elaborate rationale. Information of this sort is quite difficult for the student of expertise to use and, in addition, may not accurately represent what the expert is really doing.

Only in the past 20 yr. has there been any extensive study of the judgment process, and this study has been primarily within the context of clinical psychology. The earliest research efforts focused on the accuracy of judgments and the degree to which experts

agreed with one another in their evaluations. The results of these studies have indicated a distressing lack of accuracy and interjudge agreement both in medicine (Garland, 1959, 1960) and in clinical psychology (Goldberg, 1968).

As a result of these findings, the emphasis has shifted from research on the validity and reliability of judgments to attempts to understand the judgment process itself. This recent research aims to "simulate" or "model" the hidden cognitive processes of the judge. Hopefully, by understanding these processes it will be learned why some judges are more accurate than others, and this knowledge will, in turn, help us to train persons to make better judgments.

Some of the first models for describing quantitatively the judgment process were developed by Hoffman (1960) and by Hammond and his associates (Hammond, Hursch, & Todd, 1964). While their techniques have been quite successful in describing how individual items of information are weighted and combined by a judge, they have not been successful in describing complex patterned or configural use of information, that is, the process whereby an item of information is interpreted differently from one time to the next, depending on the nature of other available information. Since experts generally claim that they use information configurally, it is important that techniques used to describe judgment be sensitive to such processes.

One technique that analyzes the judgment process in all its complexity has been described by Kleinmuntz (1968), who had clinical psychologists and neurologists "think aloud" into a tape recorder as they made diagnostic judgments. Kleinmuntz utilized these rich introspective reports to construct a computer program simulating the diagnosticians' thought processes. The resulting programs were complex sequential (e.g., hierarchical or "tree") representations of the diagnosticians' verbal reports. At the present time it is not clear whether the failure of investigators other than Kleinmuntz to find experimental evidence for configurality stems from lack of configurality in the processes themselves or from deficiencies in the models

and procedures employed to evaluate those processes (Goldberg, 1968).

Hoffman, Slovic, and Rorer (1968) introduced a technique based on the analysis of variance (ANOVA) for the quantitative description of both configural and nonconfigural use of information in judgment. They employed this technique to study the processes whereby radiologists diagnose the malignancy of gastric ulcers on the basis of roentgenological signs. Although the radiologists were found to process information configurally in many instances, the overall influence of such nonlinear processing was slight. Most of the variability in the diagnoses could be predicted from a linear combination of signs.

Because the ANOVA technique proved quite capable of describing the use of information by individual radiologists and because it was sensitive to configural processing it appeared to merit further use. The purpose of the present paper was to test the adequacy of the ANOVA technique for describing the way that a stockbroker employs information as he evaluates the attractiveness of a company's stock. The stock market was selected as the domain in which to study expertise for several reasons. First, the task of predicting the future market price of a security is an important one. Hundreds of thousands of decisions, involving many millions of dollars, are made daily in the market. Secondly, this task is interesting because it is extremely difficult and complex. There are hundreds of factors which may be relevant, some of them economic, some of them financial, and some of them psychological in nature. In addition, introspective reports by financial analysts indicate that they believe that the relevant factors should be interpreted in a complex configural manner. For example, many analysts claim that one cannot interpret recent price changes of a stock without taking into account the volume of sales that accompanied those changes.³

METHOD

Subjects. The Ss were two young brokers. Each had about 3 yr. experience with a prominent broker-

³ Since writing this article, two very relevant references have been called to the author's attention. These are Clarkson (1962) and Anderson (1969).

age firm. While these brokers may, on occasion, merely fill a client's order, they frequently are called upon for advice, and in some instances have complete responsibility for managing a client's portfolio. These men are quite concerned about their ability to judge stocks and spend several hours each day studying the market, attempting to glean information from a variety of sources such as newspapers, the ticker tape, company reports, financial analysts' reports, etc.

Procedure. The application of ANOVA to the study of judgment is simple and direct; first one selects a set of presumably relevant factors (i.e., items of information or dimensions along which a stimulus can be described) and then he constructs stimuli such that all possible combinations of these factors are represented. When the judgments that are made about each of these stimuli are analyzed in terms of an ANOVA model, a significant main effect for Factor 1 indicates that the judge's responses varied systematically with Factor 1 independent of the levels of the other factors. This implies that Factor 1 was important to the judge. A significant interaction between Factors 1 and 2 implies that the judge was interpreting particular patterns of these factors in a configural manner; that is, the interpretation of Factor 1 differed as a function of the value taken by Factor 2.

The present task was constructed with the assistance of Broker A. When asked to list the minimum number of factors upon which he could comfortably base a recommendation about a stock, Broker A selected 11 variables commonly provided in Standard & Poor's reference reports. These variables were (a) Yield (YLD). The cash dividend income for the past year as a percentage of the market price. (b) Near Term Prospects (NTP). A one- or two-paragraph forecast concerning sales, profits, dividends, earnings, etc., for the coming year. Included is pertinent information about new products, political or economic factors bearing on the company's future, etc. (c) Earnings Quarterly Trend (EQT). A comparison of quarterly earnings over the past 4-5 yr. (d) Past Year's Performance (PYP). A synopsis of relevant statistics for the past year. Includes revenues, earnings and dividends, and political and economic factors that influenced them. (e) Profit Margin Trend (PMT). A yearly comparison indicating the trend in percentage of profit from company operations per sales dollar. Presumably this relates to the efficiency with which the company is managed and has implications for future earnings. (f) Earnings Share Yearly Trend (EYT). (g) Price/Earning Ratio (PER). The ratio of market price to net earnings per share over the past 12 mo. (h) Shares Outstanding (SO). The number of shares of common stock issued by the company. (i) Resistance Trend (RES). Trend of a line connecting several recent high points on the chart of daily price action. (j) Support Trend (SUPP). Trend of a line connecting several recent low points on the daily price chart. (k) Sales Volume Trend (VOL). Trend of the

| | | |
|--|---|---------------------------------------|
| COMPANY NUMBER _____ | | |
| Yield -----high----- | | |
| Resistance Level -----down----- | Prospects Near Term -----good----- | |
| Support Level -----up----- | | |
| Volume Trend -----up----- | | |
| Past Year's Performance -----good----- | | |
| Common Share Earnings Quarterly Trend -----up----- | | |
| INCOME STATISTICS | | |
| Profit Margin Trend -----down----- | Earnings/Share Yearly Trend -----up----- | PE Ratio Comparison -----good----- |
| CAPITALIZATION Common Stock Outstanding -----few shares----- | | |

FIG. 1. Example of a typical stimulus company.

number of shares traded per day over a recent period of time.

Next, Broker A was asked whether, in the interests of simplification, he could still make a reasonable evaluation of a company's stock if information about the 11 factors were presented in dichotomous form (e.g., yield being described as either high or low, trends as either up or down, etc.). The broker said that he could. Further questioning indicated that there would be no combination of these factors so unreasonable as to make the company seem unreal and, therefore, impossible to judge.

The next step involved the construction of hypothetical companies. Ideally it would have been desirable to combine the 11 dichotomous factors in all possible ways, but in this case that would have resulted in 2^{11} or 2048 companies, clearly an unmanageable number to judge. However, if one is willing to assume that the higher order interactions are negligible, it is possible, by means of a fractional replication design (Cochran & Cox, 1957), to evaluate the main effects and lower order interactions with a considerably reduced number of stimuli.

Previous work on judgment (Goldberg, 1968) suggested that the assumption that higher order interactions would be negligible was not too unreasonable. Therefore, hypothetical companies were constructed by combining the levels of the 11 factors according to a $1/16$ fractional replication of a 2^{11} factorial ANOVA design. This produced a set of 128 companies. This reduction of stimuli results in the confounding of main effects and two-way interactions with certain of the higher order interactions. Other high order interactions serve to estimate the error term in the ANOVA. Thus, if configural use of three

TABLE 1
RELATIVE IMPORTANCE OF THE 11 FACTORS AND THEIR SIGNIFICANT
INTERACTIONS FOR BROKER A

| Factor | Description of levels | | Mean judgment | | Magni- tude of effect ^a | MS | % Variance (ω^2) |
|---|-----------------------|---------|--|---------|--|---------|---------------------------------|
| | Level 1 | Level 2 | Level 1 | Level 2 | | | |
| Main effects | | | | | | | |
| Yield (YLD) | Low | High | 5.56 | 5.67 | .11 | .4 | .001 |
| Near Term Prospects (NTP) | Poor | Good | 4.53 | 6.70 | 2.17 | 151.0** | .334 |
| Earnings Quarterly Trend (EQT) | Down | Up | 4.87 | 6.36 | 1.49 | 70.5** | .156 |
| Past Year's Performance (PYP) | Poor | Good | 5.40 | 5.83 | .43 | 5.7* | .013 |
| Profit Margin Trend (PMT) | Down | Up | 5.44 | 5.80 | .36 | 4.1* | .009 |
| Earnings Yearly Trend (EYT) | Down | Up | 5.56 | 5.67 | .11 | .4 | .001 |
| Price/Earnings Ratio (PER) | Poor | Good | 4.80 | 6.44 | 1.64 | 86.1** | .190 |
| Shares Outstanding (SO) | Few | Many | 5.70 | 5.53 | .17 | 1.0 | .002 |
| Resistance Trend (RES) | Down | Up | 5.50 | 5.73 | .23 | 1.8 | .004 |
| Support Trend (SUPP) | Down | Up | 5.39 | 5.84 | .45 | 6.6** | .015 |
| Sales Volume Trend (VOL) | Down | Up | 5.69 | 5.55 | .14 | .6 | .001 |
| Interactions | | | | | | | |
| EYT \times SUPP | | | | | .36 | 4.1* | .009 |
| YLD \times PMT | | | | | .39 | 4.9* | .011 |
| RES \times SUPP | | | | | .48 | 7.5** | .017 |
| RES \times VOL | | | | | .36 | 4.1* | .009 |
| SUPP \times VOL | | | | | .39 | 4.9* | .011 |
| RES \times SUPP \times VOL | | | | | .42 | 5.7* | .013 |
| error | | | | | | 1.0 | |
| Sum of effects over the statistically significant factors | | | (main effects) 6.54 (73%) (interactions) 2.40 (27%) | | | | .717 .070 |
| | | | 8.94 | | | | .787 |

^a Based on the degree to which the mean judgment changes as the factor changes.
* $p < .05$.
** $p < .01$.

or more factors does occur, the error term will be inflated.

Figure 1 illustrates the way in which information about a company was displayed to the brokers. The spatial format of the variables was designed to approximate the layout of a Standard & Poor's report as closely as possible. The stimuli were bound in a notebook which the brokers took home. The brokers worked on the judgments in their leisure time over a 3-wk. period. Broker A reported spending 10½ hr. making his judgments. Broker B spent about 9 hr. at the task. Although they knew the companies were hypothetical, both brokers reported that the task was extremely interesting to them and that they were able to conjure up images of real companies as they read the stimulus information.

The brokers were asked to make a recommendation about each company based on their judgment of the likelihood that the market price of that company's stock would increase substantially in the next 6–12 mo. The recommendation was made on a 9-category rating scale where Category 1 was labeled "strong recommendation not to buy," Category 4 was a "slight recommendation not to buy," Category

5 was a "neutral" evaluation, and Categories 6 and 9 were labeled slight and strong "recommendations to buy," respectively.

RESULTS

The mean rating given the 128 companies by Broker A was 5.62 with a standard deviation of 1.94. Broker B was less favorably inclined towards the companies' stocks ($M = 3.96$) and more variable in his ratings ($SD = 2.96$).

Despite the fact that Broker B was recruited as an S by Broker A on the grounds that his approach to selecting stocks was relatively similar to that of Broker A, there was rather poor agreement between the two with regard to their ratings. The correlation between the two brokers' judgments across the 128 companies was only .32.

TABLE 2
RELATIVE IMPORTANCE OF THE 11 FACTORS AND THEIR SIGNIFICANT
INTERACTIONS FOR BROKER B

| Factor | Mean judgment | | Magnitude of effect ^a | MS | % Variance (ω^2) |
|---|---------------|---------|--|---------|---------------------------|
| | Level 1 | Level 2 | | | |
| Main effects | | | | | |
| Yield (YLD) | 3.91 | 4.02 | .11 | .4 | .000 |
| Near Term Prospects (NTP) | 3.41 | 4.52 | 1.11 | 40.0** | .082 |
| Earnings Quarterly Trend (EQT) | 3.75 | 4.17 | .42 | 5.7** | .011 |
| Past Year's Performance (PYP) | 3.89 | 4.03 | .14 | .6 | .000 |
| Profit Margin Trend (PMT) | 3.26 | 4.66 | 1.40 | 61.9** | .129 |
| Earnings Yearly Trend (EYT) | 2.91 | 5.02 | 2.11 | 142.4** | .299 |
| Price/Earnings Ratio (PER) | 3.12 | 4.80 | 1.68 | 89.4** | .187 |
| Shares Outstanding (SO) | 4.03 | 3.89 | .14 | .6 | .000 |
| Resistance Trend (RES) | 3.50 | 4.42 | .92 | 27.2** | .056 |
| Support Trend (SUPP) | 3.61 | 4.31 | .70 | 15.8** | .032 |
| Sales Volume Trend (VOL) | 3.83 | 4.09 | .26 | 2.3 | .003 |
| Interactions | | | | | |
| YLD \times PER | | | .30 | 2.8* | .005 |
| EYT \times PER | | | .55 | 9.6** | .019 |
| PYP \times PMT | | | .33 | 3.4* | .006 |
| RES \times VOL | | | .39 | 4.9** | .009 |
| SUPP \times VOL | | | .42 | 5.7** | .011 |
| error | | | | 0.6 | |
| Sum of effects over the statistically significant factors | | | (main effects) 8.34 (81%) (interactions) 1.99 (19%) | | .799 .050 |
| | | | 10.33 | | .849 |

^a Based on the degree to which the mean judgment changes as the factor changes

* $p < .05$.

** $p < .01$.

In order to isolate the factors influencing the recommendations, a separate ANOVA was performed on each broker's responses. Sums of squares and mean squares were computed for each of the 11 main effects (individual factors), each of the two-way interactions, and each of the few three-way interactions that were confounded only with four-way or higher order interactions. In addition, two indexes of the importance of a factor or interaction were computed for each effect. One was simply the standard calculation of the magnitude of an effect, based upon the degree to which the mean judgment shifted as the levels of a factor were varied. In this regard, the magnitude of a two-way interaction effect indicates the degree of change in the mean judgments as a function of variation in the levels of a pair of factors after the main effects have been partialled out. The

second index, called ω^2 , is a function of the squared magnitudes of effect and provides an estimate of the proportion of the total variance in the broker's judgments that could be attributed to a particular main effect or interaction (Hays, 1963).

Tables 1 and 2 present the results of the ANOVAs for the two brokers. The ratings of Broker A changed significantly with variation in the levels of each of six factors (main effects), the most influential of these being Near Term Prospects, Price/Earnings Ratio, and Earnings Quarterly Trend. In addition, six interactions were significant, one of these (Resistance Trend \times Support Trend) being the fourth strongest effect. Broker B exhibited seven significant main effects, the strongest of which were due to the Earnings Yearly Trend, Price/Earnings Ratio, and Profit

Margin Trend. In addition, five two-way interactions were significant.

Since the 11 factors studied here were specifically selected by Broker A as the most important ones from among a much larger set, the fact that his judgments were not influenced significantly by a number of these factors is especially noteworthy. During the process of selecting these factors the broker was able to give an elaborate rationale for including each one. Perhaps it was too difficult for him to use all of the factors simultaneously.

Summing the ω^2 index over the statistically significant factors indicated that about 72% of the variance in Broker A's responses was predictable from knowledge of six main effects and an additional 7% could be attributed to configural use of cues (significant interactions). Comparable figures for Broker B were 80% (main effects) and 5% (interactions). These percentages could be interpreted as evidence for the negligibility of configural cue utilization as were the comparable percentages found in the study of radiologists by Hoffman, Slovic, and Rorer (1968). However, the use of variance percentages as descriptive indicators may be more meaningful statistically than psychologically, and the magnitude of effect index, based upon the influence of a factor upon the mean judgments, might well be a more appropriate gauge for assessing

the relative importance of configural effects. This index indicates that configurality was substantial, accounting for 27% of the total effects on Broker A and 19% of the effects on Broker B. Even this is a conservative estimate of the degree of configurality. Extrapolating from the excellent discussions of linear and configural models presented by Green (1968) and Hayes (1968), one could argue that whenever the interaction between two factors was significant, those factors were being used configurally and the variance accounted for by both their main effects and their interaction should be counted as configural variance. Following this rule would boost the percentage of configural variance to 36% for Broker A and 85% for Broker B. Additional evidence for the argument that meaningful configural information processing was taking place here is the fact that two interactions ($RES \times VOL$ and $SUPP \times VOL$) were common to both brokers. Detailed analysis of these interactions showed each of them to be almost identical in form for the two brokers.

An index of the overall importance of a given factor was calculated by summing the magnitude of effect index for the main effect of that factor with the magnitude of effect indexes of all significant interactions containing that factor. The summed effect of a given factor was divided by the sum of the effects of all factors. This index of importance was thus a percentage score where the sum of all percentages totaled 100.

Figure 2 illustrates the relative importance of the 11 factors for each broker based on the index just described. Despite the fact that the brokers viewed themselves as similar in orientation, there was a considerable difference in their use of information. These differences undoubtedly indicate why they disagreed so often in their rating of a particular stock. Broker A considers himself to be a "technical analyst" (i.e., one who weights information from price and volume charts especially heavily) and in this regard it is noteworthy that the ANOVA model showed him to be using the three chart variables, Resistance, Support, and Volume Trends, to a greater extent than did Broker B, who ap-

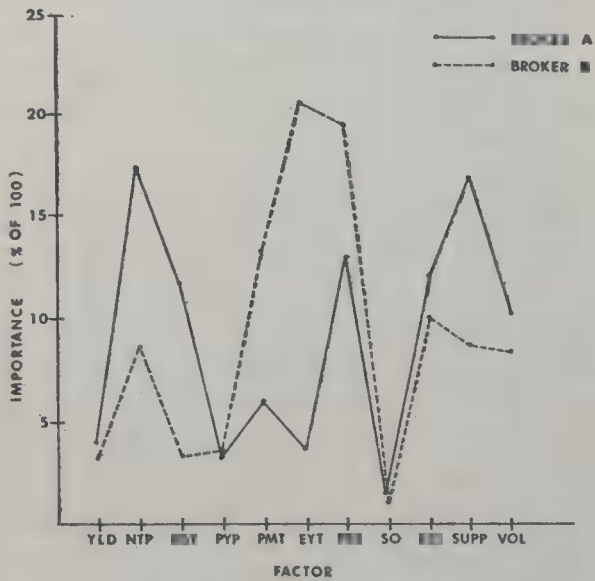


FIG. 2. The relative importance of each factor for Brokers A and B.

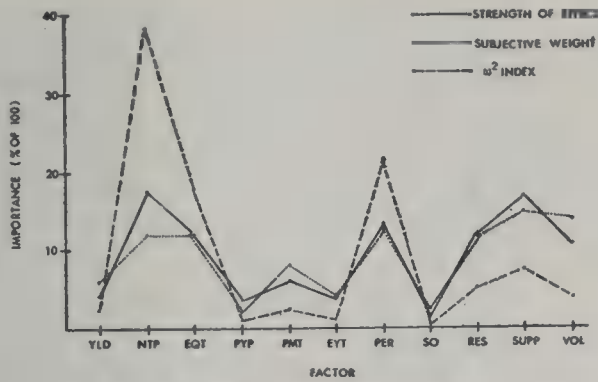


FIG. 3. Comparison between the strength of effect index, subjective weight, and ω^2 for Broker A.

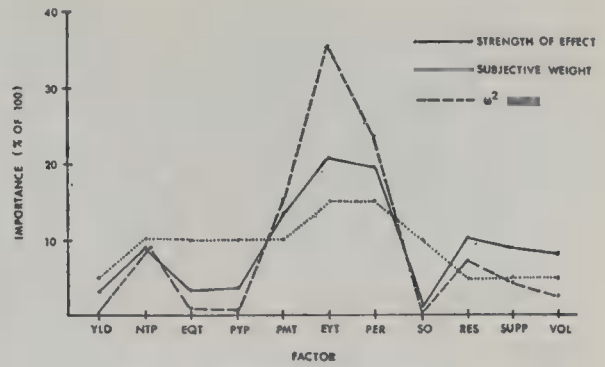


FIG. 4. Comparison between the strength of effect index, subjective weight, and ω^2 for Broker B.

appears to be more of a "fundamentalist" (i.e., one who relies on traditional balance sheet and income indicators).

Validity of subjective weights. How closely would the brokers' subjective impressions of the relative importance of the 11 factors conform to the importance indexes calculated from the ANOVA model? To provide an answer to this question, each broker was asked, after completing his ratings, to distribute 100 points over the 11 factors proportionally to his feelings about their importance in determining his judgments. These subjective weightings were compared with the magnitude of effect indexes pictured in Figure 2 and with the ω^2 index, the latter also being combined over both main effects and interactions and normed to sum to 100 over the 11 factors. The results of this comparison are depicted in Figures 3 and 4. They show that the subjective weightings of Broker A were extremely close to the magnitude of effect index while Broker B had less accurate insight into his use of the various factors. The ω^2 index was very discrepant from the subjective weights of both brokers. This index tended to exaggerate the differences between the most important factors and the lesser ones. To the extent that one feels that expert judges should have some insight about their own weighting system, this result implies that the magnitude of effect index is a better measure of a factor's relative importance than the ω^2 index.

Analysis of interactions. The finding of a significant main effect or interaction is only a first step in understanding how a judge interprets information. It should be viewed as a signal that something interesting is going

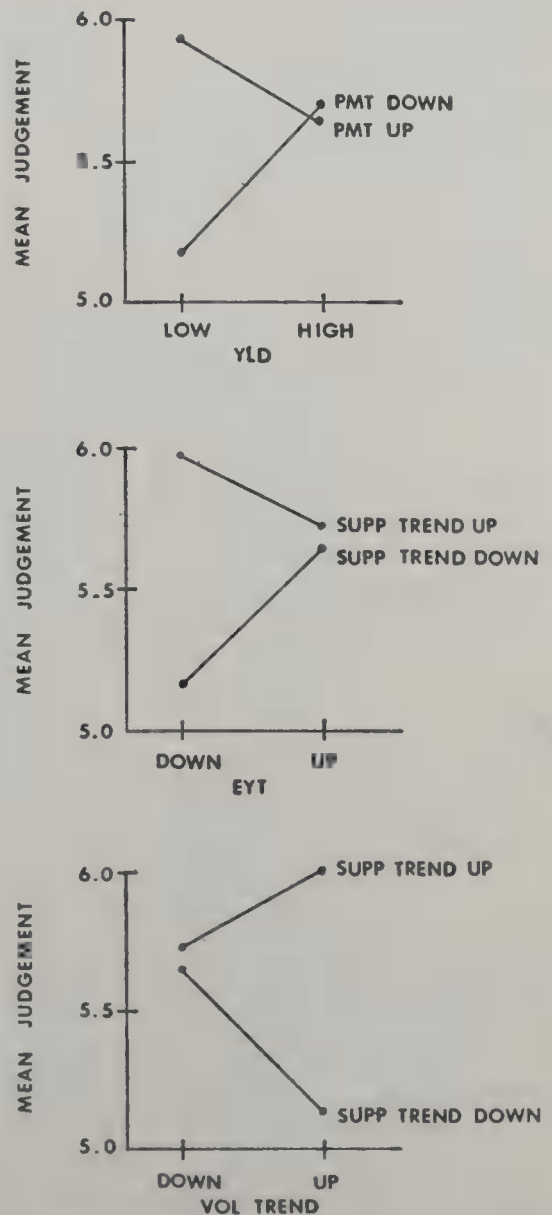


FIG. 5. Graphical representation of selected interaction effects for Broker A.

on. Graphical representation of an effect followed by interrogation of the judge concerning the rationale behind his behavior can be used to further one's understanding of the effect. To illustrate, three of the significant interactions found in the judgments of Broker A are pictured in Figure 5. Broker A was shown these figures and was asked to provide an explanation for each one. A paraphrased version of his explanation for each effect follows.

1. $YLD \times PMT$ effect. Why is high yield a more favorable indicator than low yield when PMT is down while the reverse is true when PMT is up?—Because when PMT is down, earnings probably are down, and accordingly the price of the stock should decline. A low dividend yield would make the stock even less attractive while a high yield would tend to compensate for the poor earnings prognosis. When PMT is up, earnings are probably up and the outlook for price appreciation is good. A quality company whose earnings portend good growth doesn't usually offer a large dividend, so low yield in conjunction with a rising PMT suggests that the stock has a very promising future. A high yield in this case suggests that the company is probably not putting enough of its capital into growth or perhaps that the outlook for future price appreciation is not really so promising, hence the need for a larger dividend to make the stock attractive to the investor.

2. $EYT \times SUPP$ effect. Why should a rising trend in yearly earnings be a better sign than a declining earnings trend when the support trend (price) is down while the reverse is true when the support trend is up?—When both support and earnings trends are down, the stock has nothing going for it. But if the support trend is down despite the fact that the earnings are going up, the market may be generally bad and this may be a good time to buy the stock. In contrast, when the support and earnings trends are both rising, the stock may have already made its move and thus may be overpriced, while a rising support trend in conjunction with declining earnings may indicate that the smart money knows the earnings will be up next year and the stock may be a very good buy.

3. $VOL \times SUPP$ effect. Why is rising volume viewed as a favorable indicator when it occurs in conjunction with a rising support level and viewed as a relatively unfavorable sign when it occurs with a stock whose price is declining?—A stock that is declining on relatively low volume is considered to be strong. People have enough confidence in it to hang on to it, rather than sell. If price declines on high volume the story is different. Everyone is selling and the prospects are thought to be very poor. Similarly, if volume is down on a stock that has been appreciating in price, confidence in that stock's future must be low, in contrast with a stock that is rising because many people are buying it (high volume).

DISCUSSION

The results of the present study indicate that the ANOVA technique has considerable promise as a device for describing and furthering the understanding of complex judgment processes. It is likely that this technique can provide even the expert with new insight into his inferential processes. Furthermore it might also be a valuable teaching device that would enable "trainees" to see exactly how their own processes differ from that of their expert model (see Todd & Hammond, 1965, for a related idea). Imagine the difficulty of asking the expert to describe his judgment processes in detail, obtaining a series of descriptive paragraphs such as those given above to describe interactions, and then trying to fit all these together in a way that would enable you to emulate his judgments. The task is extremely difficult if not impossible, yet this is a common way in which expertise is communicated. However, such introspective comments become considerably more helpful when they are accompanied by the precise quantitative descriptions provided by the ANOVA technique.

The present results are important in another way. They provide experimental evidence to support the commonly believed notion that judges use information configurally. The results of previous studies, most of which used less direct methods to infer the importance of configural processes, have led ■

number of workers to assert that humans are predominantly linear information processors (see discussions of this issue by Hoffman, 1968, and Goldberg, 1968). It is now clear that substantial configural processing of information does occur and can readily be detected by the ANOVA technique.

REFERENCES

- ANDERSON, N. H. Comment on "an analysis-of-variance for the assessment of configural cue utilization in clinical judgment." *Psychological Bulletin*, 1969, **72**, 63-65.
- CLARKSON, G. P. E. *Portfolio selection: A simulation of trust investment*. Englewood Cliffs, N. J.: Prentice-Hall, 1962.
- COCHRAN, W. G., & COX, G. M. *Experimental designs*. (2nd ed.) New York: Wiley, 1957.
- GARLAND, L. H. Studies on the accuracy of diagnostic procedures. *American Journal of Roentgenology, Radium Therapy, and Nuclear Medicine*, 1959, **82**, 25-38.
- GARLAND, L. H. The problem of observer error. *Bulletin of the New York Academy of Medicine*, 1960, **36**, 569-584.
- GOLDBERG, L. R. Simple models or simple processes?: Some research on clinical judgments. *American Psychologist*, 1968, **23**, 483-496.
- GREEN, B. F. Descriptions and explanations: A comment on papers by Hoffman and Edwards. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley, 1968.
- HAMMOND, K. R., HURSCHE, C. J., & TODD, F. J. Analyzing the components of clinical inference. *Psychological Review*, 1964, **71**, 438-456.
- HAYES, J. R. Strategies in judgmental research. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley, 1968.
- HAYS, W. L. *Statistics for psychologists*. New York: Holt, Rinehart, & Winston, 1963.
- HOFFMAN, P. J. The paramorphic representation of clinical judgment. *Psychological Bulletin*, 1960, **57**, 116-131.
- HOFFMAN, P. J. Cue-consistency and configurality in human judgment. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley, 1968.
- HOFFMAN, P. J., SLOVIC, P., & RORER, L. G. An analysis of variance model for the assessment of configural cue utilization in clinical judgment. *Psychological Bulletin*, 1968, **69**, 338-349.
- KLEINMUNTZ, B. The processing of clinical information by man and machine. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley, 1968.
- LUSTED, L. B. Logical analysis and roentgen diagnosis. *Radiology*, 1960, **74**, 178-193.
- TODD, F. J., & HAMMOND, K. R. Differential feedback in two multiple-cue probability learning tasks. *Behavioral Science*, 1965, **10**, 429-435.

(Received June 24, 1968)

COMMUNICATION LENGTH AS AN INDEX OF COMMUNICATOR ATTITUDE:

A REPLICATION¹

ARTHUR N. WIENS, RUSSELL H. JACKSON, THOMAS S. MANAUGH,
AND JOSEPH D. MATARAZZO²

University of Oregon Medical School

This paper attempts a replication of a study by Mehrabian (1965) who extended studies of speech behavior to apply to letters of recommendation. His Ss were asked to write two letters, both positive. For the first letter they were asked to assume strong liking for the person they were describing; for the second letter they were to assume strong dislike for the person being described. Mehrabian's Ss wrote significantly more words in the first letters. Our replication fully confirms his results. These findings suggest that the written channel of communication may be as sensitive a mirror of S's underlying attitudinal state as earlier research had revealed was the case with the spoken channel of communication.

Earlier studies reviewed in Matarazzo, Wiens, and Saslow (1965, pp. 203-204) have indicated that such interviewer tactics as head-nodding, saying Mm-Hmm, increasing his own utterance lengths, and related social reinforcers are perceived by an interviewee as "greater interest in the interviewee on the part of the interviewer." This perception, in turn, increases S's "level of satisfaction in this interpersonal encounter." This resulting motivational state was hypothesized as a preliminary conceptual framework within which to understand our often cross-validated finding that such disparate interviewer tactics were empirically, and predictably, followed by marked increases in the interviewee's own average length of utterance. In an interesting extension of these findings, Mehrabian (1965) tested for related phenomena of a communicator's underlying state in a study involving the written, in contrast to the spoken, channel of communication. He reasoned that number

of words written in a letter of recommendation might be more explicit as an index of a writer's true attitude toward the person about whom he was writing than was the content of the letter per se. His results, based on actual letters written by 69 college student Ss under two instruction-induced attitudinal sets (Like versus Dislike for the person being recommended) clearly revealed that a communicator writes more words about a person whom he likes.

In the past several years continuing research on interviewee speech behavior likewise has shifted its emphasis toward more explicit and direct concern with an S's underlying motivational and attitudinal state. Early results from this new research direction (Manauh, Wiens, & Matarazzo, in press; Matarazzo, Wiens, Manauh, & Jackson, in press) strongly suggest that discussion of some content areas elicits evidence of motivational states that are more salient than are those tapped by still other content areas. Whether such a motivationally salient state is endogenously present or exogenously induced in the interview, our own accumulating evidence is that speech behavior during an interview appears to mirror it successfully. The publication of Mehrabian's finding, then, is of considerable interest because it suggests that investigators may now be able to use the written channel of communication as well

¹ This research was carried out in conjunction with support from the Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant Number XG-3057. Thomas S. Manauh received National Institutes of Health fellowship support from NIGMS Training Grant Number GM 1495-02.

² Requests for reprints should be sent to Arthur N. Wiens or Joseph D. Matarazzo, Department of Medical Psychology, University of Oregon Medical School, 3181 S.W. Sam Jackson Park Road, Portland, Oregon 97201.

as the spoken channel in further research on underlying attitudinal and motivational states.

With Mehrabian's help the present study is an exact replication (in Portland) of his former (Los Angeles) study. The reader will find the method adequately described in Mehrabian (1965). He used college student Ss, as did the present authors (although our Ss were slightly older, upper-division collegians who were enrolled in one of two similar classes from a single instructor). In both Mehrabian's and the present study, all the data were collected in a single classroom session. For half his *N* the instructions were general (open-ended) and required each *S*, in counter-balanced order across two groups, to write two letters of positive recommendation: one about a person he "liked," the other about a person he "disliked." For the remaining half of his total *N*, Mehrabian (and the present authors) gave the writer of the two letters the additional specific information that the person *S* was writing about: "_____ is applying for a job in which you are to discuss his (her) character, intelligence, ability and perseverance at work [p. 520]." This present study used a total of 72 Ss, with *N*s of 36 and 36 in the general and specific instruction conditions, whereas Mehrabian had a total of 69, with *N*s of 38 and 31, respectively.

RESULTS

Table 1 presents the present authors' findings and includes Mehrabian's for comparison. It is clear that under either general (121.7 vs. 90.9) or specific (128.9 vs. 102.4) instructions, the present Ss also wrote more words per letter about a person they liked than they did about a person they disliked ($p < .001$ in both instances). Word counts were made by two secretaries who were unaware of the hypothesis. The Pearson r between their independent word counts was .998.

A number of additional analyses were conducted. Reading down Table 1, in neither the Like condition nor the Dislike condition did the difference between number of words written under general instructions differ significantly from the number written under specific instructions. Reading across but not shown in Table 1 (as one would expect) the number of words written by the 36 Ss (in the

TABLE 1
MEAN NUMBER OF WORDS WRITTEN IN A LETTER OF RECOMMENDATION BY SUBJECTS UNDER TWO INSTRUCTION-INDUCED ATTITUDINAL STATES

| Study | Assumed attitudinal set toward the person about whom Ss were asked to write a favorably toned letter | | |
|-------------------------------------|--|-----------------|----------|
| | Person liked | Person disliked | <i>p</i> |
| Present study (<i>N</i> = 72) | | | |
| Writers given general instructions | | | |
| <i>M</i> (No. of words) | 121.7 | 90.9 | .001 |
| <i>SD</i> | 35.5 | 29.4 | |
| Writers given specific instructions | | | |
| <i>M</i> (No. of words) | 128.9 | 102.4 | .001 |
| <i>SD</i> | 38.9 | 32.2 | |
| <i>p</i> | <i>ns</i> | | |
| Mehrabian study (<i>N</i> = 69) | | | |
| Writers given general instructions | | | |
| <i>M</i> (No. of words) | 106.4 | 92.7 | .05 |
| Writers given specific instructions | | | |
| <i>M</i> (No. of words) | 131.3 | 108.6 | .01 |
| <i>p</i> | .05 | | |

general instruction group) under the Like condition correlated highly with the number of words written by these same 36 Ss in the Dislike condition (r of .60). For the specific instruction group the comparable r was .62. Thus how many words any single *S* wrote in a letter had a fairly high degree of stability across the two attitudinal conditions. This gives even additional strength to the main finding in Table 1 that, this fact notwithstanding, taking an *S* as his own control, the average *S* in this study wrote more words when he was asked to write a positive letter with an underlying positive attitude than when he was asked to write a positive letter with an underlying negative attitude. The former attitudinal set evoked some 25% greater productivity in the two conditions (128.9 vs. 102.4 words; and 121.7 vs. 90.9 words; both $p < .001$). In still additional analyses we found (*a*) no significant effect due to the order or sequence in which the positive letters about the liked and disliked persons were written and (*b*) no correlation between an *S*'s vocabulary level as measured by the Shipley-Hartford test and the number of words written by him in any of the four conditions shown in the top half of Table 1. This latter finding strongly affirms that the presumed attitudinal effect shown in Table 1 is, in fact, that—attitudinal—and is not an effect due to intellectual competence or other

related cognitive factors which might be differentially affected by one's level of intelligence. It may be of interest to the reader that numerous attempts in our laboratory over the past decade to find differences in the noncontent speech and silence behavior (means per utterance) of Ss with very high, average, and below average WAIS IQs have failed to reveal any such differences. That is, an S's mean length of speech per utterance and his characteristic (mean) reaction time before speaking are not related to his WAIS IQ, including Verbal or Performance IQ. Thus, this similar finding with the Shipley-Hartford test in the present study adds still further support to the basic (motivational) saliency hypothesis currently being explored by the present authors, Mehrabian, and a host of other investigators.

DISCUSSION

The written and spoken channels of communication have a complex and far from consistently positive relationship with each other, for example as reported by Drieman (1962a, 1962b) from Holland in 8 Ss whose written and spoken productions were exhaustively analyzed. Nevertheless, the purpose of the present replication and modest extension of Mehrabian's study was for it to serve as a beginning bridge and test, in a second and not necessarily related channel of communica-

tion, of the fairly consistent finding with the spoken channel alone: namely, that one's underlying attitudinal or motivational state can be reflected by changes or differences in this (speech) communicative channel. The results in Table 1 suggest that the written channel of communication may be as effective as the spoken in mirroring underlying mood and other motivational states.

REFERENCES

DRIEMAN, G. H. J. Differences between written and spoken language, an exploratory study: Qualitative approach. *Acta Psychologica*, 1962, 10, 78-100. (a)
DRIEMAN, G. H. J. Differences between written and spoken language, an exploratory study: Quantitative approach. *Acta Psychologica*, 1962, 10, 36-57. (b)
MANAUGH, T. S., WIENS, A. N., & MATARAZZO, J. D. Content saliency and interviewee speech behavior. *Journal of Clinical Psychology*, in press.
MATARAZZO, J. D., WIENS, A. N., MANAUGH, T. S., & JACKSON, R. H. Interviewee speech behavior under conditions of endogenously-present and exogenously-induced motivational states. *Journal of Clinical Psychology*, in press.
MATARAZZO, J. D., WIENS, A. N., & SASLOW, G. Studies in interview speech behavior. In L. Krasner and L. P. Ullmann (Eds.), *Research in behavior modification: New developments and implications*. New York: Holt, Rinehart & Winston, 1965.
MEHRABIAN, A. Communication length as an index of communicator attitude. *Psychological Reports*, 1965, 17, 519-522.

(Early publication received February 10, 1969)

BIOGRAPHICAL CORRELATES OF ARTISTIC AND LITERARY CREATIVITY IN ADOLESCENT GIRLS¹

ANNE ANASTASI² AND CHARLES E. SCHAEFER

Fordham University

Biographical inventory keys were developed and cross-validated in a total sample of 400 public high school girls, subdivided into creative and matched control groups in art and writing. Creative Ss were selected principally through teachers' nominations supported by creative products. Cross-validation yielded criterion correlations of .34 and .55 for art and writing keys (each $p < .001$). Final keys, comprising items that differentiated in both initial and cross-validation samples (compound $p < .05$), were used to describe the biographical correlates of creativity. Comparisons were also made with the results of a similar, earlier study of creative boys.

Research on the nature and correlates of creativity has been accumulating at an increasing rate. Golann (1963) classified the various approaches with reference to their emphasis on products, process, measurement, or personality. Methodologically, investigations differ in their use of evaluated achievement (which focuses on products) or test performance as *criteria of creativity*. The test criterion is open to criticism because of limitations of test coverage and inadequate or inconsistent validation data. For these reasons, the criterion employed in this study was evaluated achievement. More specifically, the present criterion reflected the essential conditions of creativeness proposed by MacKinnon (1962), which include (a) novelty, originality, or statistical infrequency; (b) adaptiveness to reality, involving the achievement of some reality-oriented goal, such as the solution of a scientific or aesthetic problem; and (c) sustained activity leading to the development, evaluation, and elaboration of the original idea. It is apparent that tests concentrate on the first of these conditions, largely neglecting the last two.

In the effort to identify the *correlates of creativity*, different investigators have employed aptitude and personality tests, interviews, and biographical inventories. The biographical inventory provides a standardized

group procedure for gathering information about the individual's experiential history and about relevant aspects of the psychological environment in which he developed. Insofar as environment may play a significant role in the development of creativity, the biographical inventory technique should serve a dual function: (a) prediction of subsequent creative achievement in individuals, (b) identification of environmental variables conducive to the development of creative behavior.

As predictive instruments, biographical inventories have repeatedly demonstrated satisfactory validity against complex industrial, military, and educational criteria (Freeberg, 1967; Henry, 1966). With regard to creative achievement, they have proved effective in differentiating between levels of creativity in several groups of scientific research workers. Such results have been obtained with petroleum research scientists (Morrison, Owens, Glennon, & Albright, 1962; Smith, Albright, Glennon, & Owens, 1961), with a variety of research personnel in a pharmaceutical company (Buel, 1965; Tucker, Cline, & Schmitt, 1967), with engineers (McDermid, 1965), and with psychologists and chemists (Chambers, 1964). In a series of studies of scientists in the National Aeronautics and Space Administration (NASA), Taylor, Ellison, and Tucker (1966) obtained validity coefficients in the .40s and .50s when biographical inventory keys were cross-validated against several criteria of creative achievement. It is also noteworthy that such biographical inventory keys have shown substantial validity generalization when applied to research scien-

¹ This study is part of a larger project supported by Subcontract No. 2 of the Center for Urban Education, Contract OEC-1-7-062868-3060 with the United States Office of Education.

² Requests for reprints should be sent to Anne Anastasi, Department of Psychology, Fordham University, Bronx, New York 10458.

tists in other fields (Buel, Albright, & Glennon, 1966; Cline, Tucker, & Anderson, 1966).

Investigations at the high school and college level have usually employed tests as criteria, predictors, or both. Few studies have utilized biographical inventories and still fewer have done so against a criterion of evaluated achievement. Taylor, Cooley, and Nielson (1963) applied a modified version of the biographical inventory developed on NASA scientists to high school students participating in a summer science program supported by the National Science Foundation. This biographical inventory proved to be the best overall predictor of creative research performance in these students, its validity being as high as .47 in one of the groups. Parloff and Datta (1965) compared contrasted groups of participants in the Westinghouse Science Talent Search, selected on the basis of judges' ratings of their research projects. However, these groups were compared chiefly in personality test scores, the only background items reported being father's occupation, socioeconomic level, and intactness of family. Dauw (1966) successfully differentiated between highly creative and less creative adolescents by means of a biographical inventory, but his Ss were chosen on the basis of creativity tests only.

A series of studies conducted for the National Merit Scholarship Corporation report significant relationships between biographical data and subsequent creative achievement in college (Holland & Nichols, 1964; Nichols & Holland, 1963). That the obtained relationships are often low may result in part from the highly selected nature of the samples. It is of particular interest that among the many predictors investigated—including aptitude and personality tests—the best predictor of creative achievement in college was creative achievement in the same area in high school (Holland & Astin, 1962; Nichols & Holland, 1964). Even more striking is the finding that, in a large and representative sample of college freshmen, it was the students with superior high school grades who had most often won distinction for creative achievement in high school extracurricular activities (Werts, 1966). Contrary to a prevalent view, academic aptitude was closely related to creativity, especially in scientific and literary fields.

Relevant biographical data have also been obtained in studies employing interviewing or other intensive individual assessment procedures with adults who have made creative contributions in the arts or sciences (MacKinnon, 1962; Roe, 1951a, 1951b, 1953). Similar assessment techniques were utilized by Helson (1967) with college women identified through faculty nominations and ratings of creative achievement in college. In the same series of studies, Helson (1965, 1966, 1967) gathered questionnaire data regarding childhood interests and activities as recalled by her Ss. Finally, parental characteristics have been investigated in relation to children's creativity as determined by either creativity tests or evaluated achievement (Domino, 1969; Dreyer & Wells, 1966; Helson, 1966, 1967; Weisberg & Springer, 1961). The Ss of these studies included school children, high school students, and college women.

In an earlier study by the present writers (Schaefer & Anastasi, 1968), biographical inventory keys were developed in a group of 400 high school boys against criteria of creative achievement in (a) science and (b) art or creative writing. Cross-validation yielded validity coefficients of .35 and .64 for the science and art-writing keys, respectively, both significant at the .001 level. In the present study, the same basic procedures were followed in developing biographical inventory keys for high school girls in creative art and creative writing. These two fields were chosen for further exploration because in the earlier study differentiation between creative and control groups was greater in the combined art and writing group than in the science group. Among high school girls, moreover, outstanding creative achievement in art or writing is more frequent than it is in science. As in the earlier study, a second major objective of the present investigation was to utilize the differentiating biographical inventory items in formulating a description of the antecedents and correlates of creativity in this population.

METHOD

Subjects

The Ss employed in the principal data analyses were 400 female students from seven public high

schools in greater New York.³ These schools were chosen, first, because they offer courses or programs providing opportunities for creative activities and, second, because they have outstanding records of awards, prizes, and other indications of creative student achievement in art or writing. Of the 400 Ss, 246 were seniors, 128 juniors, and 26 sophomores. The group as a whole was superior with regard to educational level of parents, slightly more than one-half of the fathers and one-third of the mothers having attended college for one or more years. While over half of the parents were born in New York City, nearly one-third were foreign-born. The most frequent national ancestries were Russian, Polish, and German, in that order; 24 Ss were Negro.

The total sample comprises four criterion groups of 100 students each, designated as follows: Creative-Art (CrA), Control-Art (CoA), Creative-Writing (CrW), and Control-Writing (CoW). For inclusion in a *creative* group, S had to meet two criteria: (a) teacher nomination on the basis of one or more creative products to be listed on a teacher nomination form—any type of visual art or creative writing was acceptable for this purpose; (b) score above a minimum cutoff on Guilford Alternate Uses and Consequences tests. The *control* Ss were enrolled in the same courses from which the creative Ss were selected and were nominated by the same teachers as having provided no evidence of creative achievement. They also scored below a maximum cutoff on the two Guilford screening tests. Within each field, creative and control groups were matched in school attended, class, and grade-point average. The 400 Ss in the four criterion groups were selected from an initial pool of 1,114 nominees in the seven schools.

It should be noted that the Guilford tests were employed only as a check on irrelevant factors that might have influenced the nomination of creative or control Ss. The scores on these tests were employed only to exclude cases, never to admit them. Moreover, the two cutoff scores were sufficiently extreme as to exclude only those students whose test performance was highly discordant with their reported achievement. In terms of available published norms, the mean scores of the creative students on the two Guilford tests are approximately equal to those of college students, while the mean scores of the control groups fall close to the ninth grade mean.

Biographical Inventory

Except for minor changes, the biographical inventory employed in this study was the same as that prepared in the earlier study of high school boys (Schaefer & Anastasi, 1968). The questions were originally formulated on the basis of hypotheses and published research findings regarding the correlates

of creativity. The 166 questions of this inventory are grouped into five sections designated as physical characteristics, family history, educational history, leisure-time activities, and miscellaneous. Most of the questions cover objective facts regarding present or past activities and experiences; some call for expressions of preference and others pertain to plans and goals.

The inventory contains some multiple-choice and checklist items; but many questions are open-ended. Even with the objective items, moreover, there is usually provision for additional unlisted responses. Although scoring and data analysis are more difficult under these conditions, these types of items yield a richer return of information and are especially appropriate in an exploratory study. All responses were coded prior to tabulation. For each question, there were several possible responses, the number being quite large for some questions. In addition, several questions yielded responses that could be classified from different viewpoints to test different hypotheses. For example, a response to "List your present hobbies" could be scored with reference to number of hobbies or type of hobbies; and hobbies could be sorted into types according to several different schemas. As a result, the 166 questions yielded a total of 3,962 "scorable items" or individual response alternatives employed in the item analysis.

Procedure

The biographical inventory, together with three tests employed in another part of the project, was administered by the same E to groups of 110–256 students during a 2-hr. session held in the school buildings outside of school hours. The Ss were paid for participating in this testing session. Identification numbers were assigned to provide anonymity, and students were assured of the confidentiality of their responses.

In the analysis of biographical inventory data, each of the four criterion groups was subdivided into two subgroups of 50, employed for development of scoring keys and cross-validation, respectively. Each pair of subgroups was equated in number of students from each school, class distribution, grade-point average, and mean score on the screening tests. For each of the 3,962 scorable items, classified as present or absent, a phi coefficient was computed against the dichotomous criterion of creative versus control. These coefficients were computed separately in art and writing criterion groups. All items with phi coefficients at the significance level of $p < .20$ or better were considered for inclusion in the *initial CrA and CrW scoring keys*. Some of these items were excluded because they duplicated other items, were checked by fewer than four Ss in either subgroup, or were inconsistent with other responses or with hypotheses and hence likely to have yielded isolated chance correlations.

In the initial scoring keys, a weight of 1 was assigned to items discriminating between the $p < .20$ and $p < .05$ levels, and a weight of 2 to items discriminating at the $p < .05$ level or better. Items with higher frequencies in the creative group received

³ The authors gratefully acknowledge the cooperation of J. Wayne Wrightstone, Assistant Superintendent, Board of Education of the City of New York, Nathan Brown, then with the Center for Urban Education, and the principals and participating teachers of the following high schools: Abraham Lincoln, Art and Design, Erasmus Hall, Forest Hills, Jamaica, Midwood, and Music and Art.

TABLE 1

Biographical Inventory Scores of Criterion Groups in Cross-validation Samples

| Score | Creative art key | | Creative writing key | |
|-------------------|------------------|-------------|----------------------|-----------------|
| | Creative art | Control art | Creative writing | Control writing |
| 141-150 | 0 | 0 | 1 | 0 |
| 131-140 | 0 | 0 | 6 | 1 |
| 121-130 | 0 | 0 | 6 | 1 |
| 111-120 | 0 | 0 | 5 | 2 |
| 101-110 | 0 | 0 | 11 | 4 |
| 91-100 | 1 | 0 | 10 | 1 |
| 81-90 | 7 | 1 | 4 | 10 |
| 71-80 | 3 | 5 | 3 | 8 |
| 61-70 | 12 | 7 | 3 | 11 |
| 51-60 | 19 | 11 | 1 | 8 |
| 41-50 | 4 | 13 | 0 | 3 |
| 31-40 | 1 | 8 | 0 | 1 |
| 21-30 | 3 | 5 | 0 | 0 |
| N | 50 | 50 | 50 | 50 |
| M | 61.26 | 50.00 | 104.40 | 76.24 |
| σ | 15.65 | 15.04 | 21.57 | 20.87 |
| Range | 24-94 | 23-89 | 54-149 | 39-136 |
| z | 3.67* | | 6.63* | |
| r _{pbis} | .34* | | .55* | |

Note.—In order to eliminate negative scores, 50 was added to each raw score. This adjustment, however, does not exclude negative scores from the total possible range, which is -87 to 248 for the CrA key and -60 to 370 for the CrW key.
* p < .001.

positive weights; those with higher frequencies in the control group received negative weights. The initial CrA and CrW scoring keys were used in scoring the biographical inventories of the corresponding creative and control Ss in the cross-validation samples. The scorers were unaware of the criterion status of Ss. The scores thus obtained were correlated with the dichotomous criterion to provide an estimate of the validity of the scoring keys.

In order to utilize all the data in the selection of items for *final scoring keys*, item analyses were carried out independently in initial and cross-validation samples and those items were selected that differentiated between creative and control groups with a compound probability of .05 or better (Baker, 1952).

RESULTS

Application of the initial CrA and CrW biographical inventory keys to the appropriate cross-validation samples yielded the data summarized in Table 1. Although there is considerable overlapping between the scores of creative and control groups, the means of both creative groups are significantly higher than those of the corresponding control groups

at the .001 level. Point-biserial correlations between biographical inventory scores and the dichotomous criterion are .34 in the art group and .55 in the writing group.

At least two conditions imposed upon the selection of Ss tend to reduce the differences between creative and control groups. First, creative and control Ss were equated in grade-point average, although there is evidence that high school grades are in fact related to creative achievement (e.g., Werts, 1966). Second, the creative and control Ss were enrolled in the same courses in art or writing and attended high schools noted for the creative achievement of their students.

The second condition applies more strongly to the art than to the writing group, since a large proportion of Ss in the art sample were in special high schools whose students are selected on the basis of superior artistic talents. This fact is consistent with the finding that differentiation between creatives and controls was less sharp in the art than in the writing group. Not only were the mean difference and the criterion correlation higher in the writing than in the art group, but the number of significantly differentiating items was also larger in the CrW key than in the CrA key—a difference that is reflected in the higher scores obtained with this key. In the light of these sample characteristics, it should be noted that the present study is concerned with the differentiating biographical characteristics of the more highly creative Ss within an academically superior and talented population.

After the cross-validation of the initial biographical inventory keys, final keys were constructed with items whose compound probability was derived from both initial and cross-validation samples. The CrA key thus developed contains 40 items, the CrW key 82 items. An examination of these items provides a description of the biographical correlates of creativity as revealed within the conditions of this study.

DISCUSSION

Correlates of Creativity across Both Fields

The most conspicuous characteristic of the creatives in both fields is a *pervasive and continuing interest* in their chosen field and ab-

sorption in its pursuit. Items in this category include those dealing with favorite subjects in elementary school and high school; subjects found easy and those found difficult; nature of extracurricular activities in elementary school and high school, as well as anticipated extracurricular activities in college; concentration of hobbies in one's field of interest, as well as hobbies bearing a close relation to vocational goal; and reported career plans. Strength of interest is also indicated by the significantly greater number of creatives than controls in both fields reporting that they frequently became so absorbed in a project that they missed a meal or stayed up late.

Typically, the highly creative adolescent girl in this study had manifested an absorbing interest in her field since childhood and her creative activities had received recognition through exhibitions, publication, prizes, or awards. Her initial interest was thus rewarded and reinforced early in life by persons in authority, such as parents and elementary school teachers. The continuity of creative achievement over time is corroborated by the findings of other investigations, notably Helson's (1965, 1967) research with college women, the surveys of National Merit Scholarship finalists (Holland & Astin, 1962; Nichols & Holland, 1964), and our own earlier study of creative high school boys (Schaefer & Anastasi, 1968).

Several significantly differentiating items suggest a predominance of *unusual experiences* in the backgrounds of the creatives as contrasted with the controls. Thus the creatives were more likely than the controls to have had a variety of unusual experiences, to daydream about unusual things, to have collections of an unusual nature (such as ant pictures, mushrooms, and mobiles), and to have experienced eidetic imagery or had imaginary companions in childhood. To some extent, these differences may indicate greater readiness to acknowledge unusual experiences on the part of the creatives and less reluctance to report them. It is also interesting to note that more creatives than controls in both fields reported unusual types of paternal discipline, other than those listed on the inventory form. One could speculate that the prevalence of

atypical experiences in their early life may contribute to the low level of conformity and conventionality generally found to characterize creative persons at all ages.

Because of the selection procedures employed, both creative and control groups tended to come from intellectually superior homes. Nevertheless, certain significant differences were found in the *familial backgrounds* of creatives and controls. In both creative groups, significantly more fathers had attended college, graduate school, or professional schools than was true in the corresponding control groups. More controls than creatives reported that no musical instrument was played in the family. Since Ss were not selected for this study on the basis of musical achievement, this difference probably reflects the general cultural level of the home. Also relevant to general home conditions may be the fact that significantly more creatives than controls reported having two or more collections.

Earlier investigations have repeatedly found creativity to be related to parental educational and occupational level and to socioeconomic level of the home, whether Ss be distinguished scientists (Chambers, 1964) or creative high school students (Schaefer & Anastasi, 1968). Nor is the relationship limited to full-fledged creative achievement. Using performance on the Minnesota Tests of Creative Thinking as a criterion, Daww (1966) found that high-scoring high school seniors had parents with better educational backgrounds and more professional and managerial occupations than did the low scorers. Similarly, in a study of seventh grade children subdivided on the basis of scores on an originality battery, socioeconomic status yielded the largest group difference of all variables investigated (Anderson & Cropley, 1966). In explaining this finding, the authors refer first to typical lower-class parental attitudes that tend to evoke anxiety toward school learning and hence encourage convergent rather than divergent thinking. As a second reason, they cite the more varied and stimulating environment provided by homes at higher socioeconomic levels. In Parloff and Datta's (1965) study of highly selected participants in the Westinghouse Science Talent Search, the

entire sample excelled above the general population in socioeconomic level and in parental educational and occupational level, although these variables were unrelated to the rated creativity of projects within the sample.

With regard to parental influence on the creative high school girls in our study, the majority of items differentiating between creatives and controls refer to the *father* rather than to the mother. In our earlier study of high school boys, the reverse was true, more of the differentiating items pertaining to the mother. These findings are consistent with those reported by Dauw (1966) for high school seniors, by MacKinnon (1962) for creative male architects, and by Helson (1966, 1967) for creative women mathematicians and creative college women. In the study of women mathematicians, moreover, Helson (1966, p. 21) reports that "the creative women were judged by interviewers to have had more identification with their fathers than comparison subjects." If such results truly indicate a greater influence of the opposite-sex parent on creative children, they may help to explain the finding that in their attitudes, interests, and problem-solving styles creative individuals show more traits of the opposite sex than do controls and generally conform less closely to sex stereotypes (see e.g., MacKinnon, 1962).

Differences between Creativity Correlates in Art and Writing Groups

The CrA and CrW groups are not directly comparable because of differences in school and class distribution and grade-point average. As might be anticipated, the grades in the CrW group average significantly higher than those in the CrA group. In the present experimental design, each creative group was equated with its own control group in these variables. The question now to be considered is whether the characteristics that significantly differentiate CrA Ss from their own controls differ in any systematic way from those that significantly differentiate the CrW Ss from their controls. This question can be answered by examining the items in the final CrA and CrW keys.

As previously noted, the CrW key contains about twice as many items as the CrA key. With few exceptions, these additional items

fall into a cluster indicative of strong intellectual and "cultural" orientation and breadth of interests, both in the student herself and in her home background. The fathers of the CrW girls, as compared with those of the controls, are more likely to have one or more hobbies, frequently of an artistic or literary nature. Magazines regularly available at home are more likely to be of the cultural-intellectual types. The student herself is more likely to own classical records, attend concerts, and read more than 10 books a year, preferably in science, science fiction, philosophy, languages, or history. She regularly reads more than two sections of a newspaper, including editorials. She frequently visits art museums and galleries, has received lessons in arts or crafts, and has a large number of hobbies, beginning in childhood, to which she now devotes over 5 hr. a week. She reports owning a microscope more often than do the controls. In high school, she participates more extensively in extracurricular activities and anticipates more participation in college. Her college plans are more fully developed and ambitious. In comparison to the controls, the CrW student is more often considering two or more colleges, usually including an Ivy League or small private college, and is less often considering a public city college.

It is noteworthy that the breadth of interests and intellectual orientation characterizing the CrW girls was found in *both* creative groups of boys in our earlier study (Schaefer & Anastasi, 1968). One of these groups was selected because of creative achievement in science, the other because of creative achievement in art or writing. The latter group, however, included 76 boys in creative writing and only 24 in art. It is thus likely that the similarity of this group to the CrW girls resulted from the predominance of creative writing cases within it.

When the results of the two studies are considered together, they indicate that the biographical correlates of creativity are closely similar for boys and girls, with the possible exception of the reversal of role model and the greater influence of the opposite-sex parent upon the creative offspring. With regard to field of creative achievement, certain characteristic differences emerge among science, writing, and art. Cutting across both sex and

field, however, are certain common characteristics of creative adolescents: continuity and pervasiveness of interest in chosen field; prevalence of unusual, novel, and diverse experiences; and educational superiority of familial background.

REFERENCES

- ANDERSON, C. C., & CROPLEY, A. J. Some correlates of originality. *Australian Journal of Psychology*, 1966, 18, 218-227.
- BAKER, P. C. Combining tests of significance in cross-validation. *Educational and Psychological Measurement*, 1952, 12, 300-306.
- BUEL, W. D. Biographical data and the identification of creative research personnel. *Journal of Applied Psychology*, 1965, 49, 318-321.
- BUEL, W. D., ALBRIGHT, L. E., & GLENNON, J. R. A note on the generality and cross-validity of personal history for identifying creative research scientists. *Journal of Applied Psychology*, 1966, 50, 217-219.
- CHAMBERS, J. A. Relating personality and biographical factors to scientific creativity. *Psychological Monographs*, 1964, 78, No. 7 (Whole No. 584).
- CLINE, V. B., TUCKER, M. F., & ANDERSON, D. R. Psychology of the scientist: XX. Cross-validation of biographical information predictor keys across diverse samples of scientists. *Psychological Reports*, 1966, 19, 951-954.
- DAUW, D. C. Life experiences of original thinkers and good elaborators. *Exceptional Children*, 1966, 32, 433-440.
- DOMINO, G. Maternal personality correlates of sons' creativity. *Journal of Consulting and Clinical Psychology*, 1969, 33, 180-183.
- DREYER, A., & WELLS, M. Parental values, parental control, and creativity in young children. *Journal of Marriage and the Family*, 1966, 28, 83-88.
- FREEBERG, N. E. The biographical information blank as a predictor of student achievement. *Psychological Reports*, 1967, 20, 911-925.
- GOLANN, S. E. Psychological study of creativity. *Psychological Bulletin*, 1963, 60, 548-565.
- HELSON, R. Childhood interest clusters related to creativity in women. *Journal of Consulting Psychology*, 1965, 29, 352-361.
- HELSON, R. Personality of women with imaginative and artistic interests: The role of masculinity, originality, and other characteristics in their creativity. *Journal of Personality*, 1966, 34, 1-25.
- HELSON, R. Personality characteristics and developmental history of creative college women. *Genetic Psychology Monographs*, 1967, 76, 205-256.
- HENRY, E. R. Conference on the use of biographical data in psychology. *American Psychologist*, 1966, 21, 247-249.
- HOLLAND, J. L., & ASTIN, A. W. The prediction of the academic, artistic, scientific, and social achievement of undergraduates of superior scholastic aptitude. *Journal of Educational Psychology*, 1962, 53, 132-143.
- HOLLAND, J. L., & NICHOLS, R. C. Prediction of academic and extracurricular achievement in college. *Journal of Educational Psychology*, 1964, 55, 55-65.
- MACKINNON, D. W. The nature and nurture of creative talent. *American Psychologist*, 1962, 17, 484-495.
- MCDERMID, C. D. Some correlates of creativity in engineering personnel. *Journal of Applied Psychology*, 1965, 49, 14-19.
- MORRISON, R. F., OWENS, W. A., GLENNON, J. R., & ALBRIGHT, L. E. Factored life history antecedents of industrial research performance. *Journal of Applied Psychology*, 1962, 46, 281-284.
- NICHOLS, R. C., & HOLLAND, J. L. Prediction of the first year college performance of high aptitude students. *Psychological Monographs*, 1963, 77, No. 7 (Whole No. 570).
- NICHOLS, R. C., & HOLLAND, J. L. The selection of high aptitude high school graduates for maximum achievement in college. *Personnel and Guidance Journal*, 1964, 43, 33-40.
- PARLOFF, M. B., & DATTA, L. E. Personality characteristics of the potentially creative scientist. *Science and Psychoanalysis*, 1965, 8, 91-106.
- ROE, A. A psychological study of eminent biologists. *Psychological Monographs*, 1951, 65, No. 14 (Whole No. 331). (a)
- ROE, A. A psychological study of physical scientists. *Genetic Psychology Monographs*, 1951, 43, 121-235. (b)
- ROE, A. A psychological study of eminent psychologists and anthropologists, and a comparison with biologists and physical scientists. *Psychological Monographs*, 1953, 67, No. 2 (Whole No. 287).
- SCHAEFER, C. E., & ANASTASI, A. A biographical inventory for identifying creativity in adolescent boys. *Journal of Applied Psychology*, 1968, 52, 42-48.
- SMITH, W. J., ALBRIGHT, L. E., GLENNON, J. R., & OWENS, W. A. The prediction of research competence and creativity from personal history. *Journal of Applied Psychology*, 1961, 45, 59-62.
- TAYLOR, C. W., COOLEY, G. N., & NIELSEN, E. C. Identifying high school students with characteristics needed in research work. NSF-G17543, University of Utah, 1963. (Mimeograph)
- TAYLOR, C. W., ELLISON, R. L., & TUCKER, M. F. *Biographical information and the prediction of multiple criteria of success in science*. Greensboro, N. C.: Richardson Foundation, 1966.
- TUCKER, M. F., CLINE, V. B., & SCHMITT, J. R. Prediction of creativity and other performance measures from biographical information among pharmaceutical scientists. *Journal of Applied Psychology*, 1967, 51, 131-138.
- WEISBERG, P. S., & SPRINGER, K. J. Environmental factors in creative function. *Archives of General Psychiatry*, 1961, 5, 555-564.
- WERTS, C. E. The many faces of intelligence. *National Merit Scholarship Corporation Research Reports*, 1966, 2, No. 5.

(Received May 5, 1968)

PERCEIVED NEED SATISFACTIONS OF DOMESTIC VERSUS OVERSEAS MANAGERS¹

JOHN M. IVANCEVICH²

Department of Administrative Science, University of Kentucky

This study compares the perceptions of need satisfactions of (a) American managers currently working in overseas locations as a function of their position in the organizational hierarchy and (b) the overseas American managers and their domestic counterparts. A need satisfaction questionnaire (Porter, 1961), previously employed in numerous studies to test the Maslow need-hierarchy concept, was used to collect the data from 127 overseas Americans. The domestic manager data used for comparisons were collected in a study by Porter (1963b). The findings indicate that no matter how Ss were classified, job level or domestic and overseas, the Autonomy and Self-actualization need categories are the least fulfilled.

Studies have been conducted recently using industrial managers, government managers (Edel, 1967; Paine, Carroll, & Leete, 1966) and union officers (Miller, 1964) as Ss in determining need satisfactions. Each of the studies used Ss who were working within the United States or who were nationals of foreign nations working in their native country. For example, a series of studies reported by Porter (1963a) found that line managers derive more need satisfaction than staff managers from their job, that the higher a manager is in the organizational hierarchy the more need satisfaction attained (Porter, 1962), and that high-level managers in large organizations have more need satisfaction than high-level managers in small firms but the reverse is true for lower level managers (Porter, 1963b).

The Porter (1961) need satisfaction questionnaire was employed to study 3,641 managers from around the world (Haire, Ghiselli, & Porter, 1966). These managers were from 14 different countries. This investigation was concerned with manager attitudes and how they were similar or different among countries. One of the general findings of this job attitude study was that over all countries included, two needs stand out as largely un-

satisfied—Autonomy and Self-actualization. The deficit in satisfaction of these two needs combined is more than twice as large as the dissatisfaction in the three other need areas, Security, Social, and Esteem.

Each of the managers studied by Haire, Ghiselli, and Porter (1966) were working within their own country. For example, the American managers included worked within the United States and the Japanese respondents were working in Japan. There are presently no empirical studies available in the current literature which compare domestic and overseas American managers at similar organizational levels on the need satisfaction derived from their job.

In an attempt to fill in some of the gaps in previous research involving managerial motivation, the present study makes use of American managers currently working overseas and study results reported by Porter (1963b) involving domestic managers. Attempts to determine if the American manager's level in the overseas managerial hierarchy influences his perceptions of need satisfaction are lacking in the literature. In addition, it has not yet been demonstrated empirically that there are differences or similarities in the need satisfaction opportunities of managers working within or outside of the United States.

If those responsible are to do the best possible job recruiting and selecting American managers for overseas assignments, it would be desirable to know what features of the overseas job are most satisfying and which

¹ The author wishes to acknowledge the cooperation of Lyman W. Porter in providing statistical data so that comparisons could be made.

² Requests for reprints should be sent to the author, Department of Administrative Science and Quantitative Methods, Commerce Building, University of Kentucky, Lexington, Kentucky 40506.

are least satisfying. For example, if it is found that the overseas job offers optimum opportunities for satisfaction of the Esteem needs of the manager when compared with a domestic position, it may be advantageous to secure overseas managerial candidates that value Esteem need satisfaction above any other.

METHOD

Sample

The study is divided into two major phases. First, data were obtained by administering Porter's (1961) need satisfaction questionnaire to top and middle managers currently working abroad representing large United States business corporations. The top management group includes presidents and vice-presidents and the middle management group consists of division, plant, and major departmental managers. Usable replies were received from 78 top managers and 49 middle managers.

The managers studied were randomly selected from name lists of American managers currently overseas working for the largest United States industrial corporations. These lists were compiled by using *Fortune's* "500," *The Directory of Firms Operating in Foreign Countries* (Angel, 1966), and names submitted by a number of consulting agencies. The United States executives included in this study have permanent assignments outside the geographical boundaries of the United States. Each of the managers have only minimal contact with their home offices. They interact primarily with the nationals of the host nation or with third country nationals.

The domestic manager need satisfaction data for the second phase which was needed to make comparisons between domestic and overseas managers were obtained from a study reported by Porter (1963b). He administered his satisfaction questionnaire to several thousand individuals in management positions in firms located throughout the continental United States. These individuals represented a random 10% sample of the American Management Association and a random sampling of a nonmember mailing list of the Association. The questionnaires were mailed to top-, middle-, and lower level managers in large-, medium-, and small-sized companies. Only the top and middle-level managers in large organizations of the Porter (1963b) study are used to make comparisons with the present investigation.

Questionnaire

The amount of need satisfaction experienced by management respondents in the Porter (1963b) study and the present study for each of 12 need items was determined by subtracting the responses to Part *a* of the item (How much is there now?) from the responses to Part *b* of the item (How much should there be?).

Individual scores were then averaged so that comparisons of the two groups on each need item could

be made. A two-tailed *t* test was used to test for statistically significant differences in perceived need satisfaction of the top and middle manager classification and the domestic and overseas manager breakdown.

Categories of Needs and Specific Item Descriptions

Listed below are the specific need items and major need categories used in both studies. The items were listed in a random fashion in the questionnaires, but are listed here in systematic order which corresponds to the Maslow-type (1943) need framework.

1. Security need
 - a. the feeling of security in my management position
2. Social needs
 - a. the *opportunity*, in my management position, to give help to other people
 - b. the *opportunity to develop close friendships* in my management position
3. Esteem needs
 - a. the feeling of *self-esteem* a person gets from being in my management position
 - b. the *prestige* of my management position inside the company (that is, the regard received from others in the company)
 - c. the *prestige* of my management position outside the company (that is, the regard received from others not in the company)
4. Autonomy needs
 - a. the *opportunity* for independent thought and action in my management position
 - b. the *authority* connected with my management position
 - c. the *opportunity*, in my management position, for participation in the *settings of goals*
5. Self-actualization needs
 - a. the *opportunity* for personal growth and development in my management position
 - b. the feeling of *self-fulfillment* a person gets from being in my management position (that is, the feeling of being able to use one's own unique capabilities, realizing one's potentialities)
 - c. the *feeling* of worthwhile accomplishment in my management position

RESULTS

Table 1 compares the need satisfaction and need category cluster scores of the top and middle manager respondents currently overseas. The individual item need satisfaction scores (e.g., 12 item scores) indicate that top managers are significantly more satisfied than middle managers with respect to their need for prestige inside the company and opportunity to participate in goal setting. The top managers report more satisfaction

TABLE 1
AVERAGE NEED SATISFACTION AND CLUSTER SCORES OF OVERSEAS AMERICAN MANAGERS STUDIED:
TOP VERSUS MIDDLE MANAGERS

| Need categories and items: | Mean scores— top managers (<i>N</i> = 78) | Mean scores— middle managers (<i>N</i> = 49) | <i>t</i> ratio |
|--|--|---|----------------|
| 1. Security need | | | |
| a. (security in job) | .413 | .653 | 1.669 |
| 2. Social needs | | | |
| a. (opportunity to help people) | .364 | .428 | .517 |
| b. (opportunity for friendships) | .310 | .489 | 1.230 |
| 3. Esteem needs | | | |
| a. (feeling self-esteem) | .240 | .408 | 1.519 |
| b. (prestige inside company) | .256 | .551 | 2.410* |
| c. (prestige outside company) | .243 | .367 | 1.176 |
| 4. Autonomy needs | | | |
| a. (opportunity for independent thought and action) | .472 | .653 | 1.748 |
| b. (authority in position) | .662 | .979 | 1.152 |
| c. (opportunity to participate in goal setting) | .405 | .775 | 2.279* |
| 5. Self-actualization needs | | | |
| a. (opportunity for personal growth and development) | .780 | .653 | — .168 |
| b. (feeling self-fulfillment) | .893 | 1.102 | 1.024 |
| c. (feeling of worthwhile accomplishment) | .770 | 1.040 | 1.737 |
| Cluster scores by rank | | | |
| Esteem | .247 | .442 | |
| Social | .338 | .459 | |
| Security | .413 | .653 | |
| Autonomy | .514 | .803 | |
| Self-actualization | .781 | .932 | |

Note.—The larger the numerical values the less perceived need satisfaction.
* *p* < .05 as determined by two-tailed *t* test.

than middle managers in 11 of the 12 need items.

The cluster scores for each category are developed so that comparisons of the overall scores can be readily made. The average mean cluster scores are presented in rank order format in Table 1. Analysis of Table 1 illustrates that the Esteem and Social needs are the most satisfied. Further review of Table 1 indicates that the largest need deficiencies for both groups are reported in the Self-actualization category. This finding generally agrees with the results (Miller, 1964; Paine, Carroll, & Leete, 1966; Porter, 1963b).

The data presented in Table 2 compares the need satisfaction and need cluster scores of domestic (Porter, 1963b) and overseas American managers. Examination of the top manager need item scores show that the overseas respondents report more satisfaction in 8 of the 12 item scores.

The middle manager scores in Table 2 indicate that the overseas manager perceives more satisfaction than his domestic counterpart on 6 of the 12 items. The largest difference in average need satisfaction occurs in comparing the 1a (security in job) scores.

The need cluster scores presented in Table 2 show that overseas top managers report more satisfaction than domestic top managers in all categories but the Social, while overseas middle managers are more satisfied than their domestic counterparts in three of the five categories. Both the domestic and overseas top and middle management groups indicate that the Autonomy and Self-actualization need categories are the least satisfied.

DISCUSSION

The results presented in Table 1 indicate that for the overseas managers included in this study there exists a relationship between

TABLE 2

AVERAGE NEED SATISFACTION AND CLUSTER SCORES OF TOP AND MIDDLE MANAGERS
OVERSEAS VERSUS TOP AND MIDDLE MANAGERS IN THE UNITED STATES

| Need categories and items | Top managers | | | Middle managers | | |
|---|--------------------------------|--------------------|----------|--------------------------------|--------------------|----------|
| | Porter— domestic N = 136 | Overseas N = 78 | t ratio | Porter— domestic N = 268 | Overseas N = 49 | t ratio |
| 1. Security needs | | | | | | |
| a. (security in job) | .36 | .41 | -.37 | .28 | .65 | -2.39* |
| 2. Social Needs | | | | | | |
| a. (opportunity to help people) | .47 | .36 | .98 | .39 | .43 | -.34 |
| b. (opportunity for friendships) | .06 | .31 | -2.31* | .25 | .49 | -1.51 |
| 3. Esteem needs | | | | | | |
| a. (feeling of self-esteem) | .63 | .24 | 3.51* | .71 | .41 | 2.47* |
| b. (prestige inside company) | .39 | .26 | 1.23 | .56 | .55 | .08 |
| c. (prestige outside company) | .23 | .24 | -.10 | .34 | .37 | -.28 |
| 4. Autonomy needs | | | | | | |
| a. (opportunity for independent thought and action) | .49 | .66 | -1.32 | .62 | .98 | -2.21* |
| b. (authority in position) | .56 | .47 | .78 | .92 | .65 | 1.63 |
| c. (opportunity to participate in goal setting) | .69 | .41 | 2.30* | .97 | .78 | 1.12 |
| 5. Self-actualization needs | | | | | | |
| a. (opportunity for growth and development) | .80 | .68 | .85 | .88 | .65 | 1.59 |
| b. (feeling of self-fulfillment) | .83 | .89 | -.38 | 1.00 | 1.10 | -.52 |
| c. (feeling of accomplishment) | .85 | .77 | .61 | 1.15 | 1.04 | .75 |
| | Domestic | | Overseas | Domestic | | Overseas |
| Security | .36 | | .43 | .28 | | .65 |
| Social | .27 | | .34 | .32 | | .46 |
| Esteem | .42 | | .25 | .54 | | .44 |
| Autonomy | .58 | | .51 | .84 | | .80 |
| Self-actualization | .83 | | .78 | 1.01 | | .93 |

Note.—The larger the numerical values the less perceived need satisfaction.

* $p < .05$ as determined by two-tailed t test.

the overseas manager's vertical level in the organization hierarchy and the opportunity to satisfy only two specific need items. The top managers reported significantly more prestige and goal setting opportunities within the company than the middle managers. In most need item scores only relatively small differences were found between the two management levels. This finding is contrary to some of the research findings of studies that investigate the relationship of perceived need satisfactions and organizational level of domestic managers (Edel, 1967; Porter, 1962).

The present study's results when compared with the findings of Porter (1963b) show some differences in opportunities to satisfy specific needs for domestic and overseas

managers. The domestic managers (Porter, 1963b), at both levels of the managerial hierarchy, reported that the need for security was highly fulfilled. It is also found that the social needs of domestic managers are generally more satisfied than their overseas counterparts. The overseas managers, however, perceived more esteem need satisfaction than domestic managers.

Because the overseas American managers are working in a foreign location away from critical scrutiny of the home office, it may be postulated that they have more autonomy than managers working in the United States. If more self-reliance exists in overseas assignments, then the American executive abroad should report at least as much satisfaction of the Autonomy need items as do the

domestic managers. Table 2 illustrates that this is not the case with respect to every need item studied.

The Autonomy and Self-actualization needs appear to be the most critical area of need fulfillment deficiency at all levels of management for both domestic and overseas American executives. This finding agrees with the Haire, Ghiselli, Mason, and Porter (1966) international study results which showed that managers throughout the world have not been able to satisfy the Autonomy and Self-actualization needs to their fullest.

REFERENCES

- ANGEL, J. *Directory of firms operating in foreign countries*. New York: World Trade Academy Press, Inc., 1966.
- EDEL, E. C. A study of managerial motivation, *Personnel Administration*, 1967, 29, 31-38.
- HAIRE, M., GHISELLI, E. E., & PORTER, L. W. *Managerial thinking: An international study*. New York: Wiley, 1966.
- MASLOW, A. H. A theory of human motivation, *Psychological Review*, 1943, 50, 370-396.
- MILLER, E. L. A study of job attitudes of national union officers. Unpublished doctoral dissertation, Department of Business Administration, University of California, 1964.
- PAINE, F. T., CARROLL, S. J., & LEETE, B. A. Need satisfactions of managerial personnel in a government agency, *Journal of Applied Psychology*, 1966, 50, 247-249.
- PORTER, L. W. A study of perceived need satisfactions in bottom and middle management jobs, *Journal of Applied Psychology*, 1961, 45, 1-10.
- PORTER, L. W. Job attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level, *Journal of Applied Psychology*, 1962, 46, 375-384.
- PORTER, L. W. Job attitudes in management: II. Perceived deficiencies in need fulfillment as a function of line versus staff type of job, *Journal of Applied Psychology*, 1963, 47, 267-275. (a)
- PORTER, L. W. Job attitudes in management: IV. Perceived deficiencies in need fulfillment as a function of size of company, *Journal of Applied Psychology*, 1963, 47, 386-397. (b)

(Received May 26, 1968)

SOURCES OF VARIATION IN JOB AND LIFE SATISFACTION: THE ROLE OF COMMUNITY AND JOB-RELATED VARIABLES

CHARLES L. HULIN¹

University of Illinois, Urbana

Several hypotheses relevant to the analysis of the effects of community characteristics on job satisfaction were tested. The Ss were 390 male and 80 female white-collar workers employed by the same company and living in two company towns in Canada. The two towns differed along certain dimensions. Predictions were made regarding the differences in reactions by the workers to these two communities. Predictions were also made regarding the relationship between responses to the communities and responses to general job and life satisfaction. Sex differences were present but the data supported the hypotheses. A discussion of the relevance of these data for job satisfaction and motivation theory is presented.

Several articles have appeared recently which analyze the role played by environmental variables in determining job satisfaction and motivation. These studies (Blood & Hulin, 1967; Hulin, 1966; Hulin & Blood, 1968; Katzell, Barrett, & Parker, 1961; Kendall, 1963; Turner & Lawrence, 1965) used traditional S-R paradigms. In some of these studies characteristics of the community in which the plant being studied was located were assessed by means of data from census tracts (to index such variates as cost of living, standard of living, slums, extent of urbanization, etc.) or state-published population figures (to index town size).

These community characteristics have been used to predict either workers' satisfaction with various aspects of the job, behaviors in the job situation, or the relationship between satisfaction and job characteristics. The results indicate that community characteristics can be used very effectively as predictors of mean responses to the job and as moderators of relationships between job characteristics and worker responses. Individual differences in preferences for work role outcomes can be predicted using such community variables (Blood & Hulin, 1967; Hulin & Blood, 1968;

Turner & Lawrence, 1965). These significant predictions of preferences for work role outcomes suggest reasons for the futility of the search for general laws of satisfaction and motivation in industrial psychology. No longer can general hypotheses be formulated which state that "Workers want larger jobs, more responsibility, more autonomy, opportunities for self-actualization, etc."

Of equal interest are the predictions of satisfaction levels from selected community characteristics (Hulin, 1966; Kendall, 1963). These latter two studies yielded results consistent with the frame of reference or adaptation level hypothesis of job satisfaction which suggests that the workers' responses to various levels of environmental return are influenced as much by what they see other workers getting as by what they get (Smith, Kendall, & Hulin, 1969). Community characteristics were chosen which were theoretically related to the frame of reference of the individual workers for judging the quality of any given level of environmental return. That is, cost of living and standard of living in the communities were chosen to index the workers' economic frame of reference. The results confirmed the economic frame of reference hypothesis in that workers in communities with high costs and standards of living were less satisfied with their pay even though in Hulin's study pay was constant across communities. The general effects of community characteristics on job attitudes are also demonstrated.

¹ The author would like to thank H. Peter Dachler, Linda Yarham, and the officials of the company involved for their help and cooperation in the research described in this paper and George Graen and Harry Triandis who read and commented on an earlier draft. Requests for reprints should be sent to the author, Department of Psychology, University of Illinois, Urbana, Illinois 61801.

The assumption made in all of these studies is that community characteristics are indexes of a certain class of psychological variables. The economic characteristics of the communities (cost of living, standard of living, etc.) do not have a direct effect on the pay satisfaction of the workers in the community. Rather, it is assumed that the effects of the economic standards of the community are mediated through intervening psychological variables (frames of reference, adaptation levels) to produce their impact on pay satisfaction. However, this assumption has not been tested in any of the research published to date. The actuarial relationship between environmental characteristics on the one hand and job attitudes on the other is all that has been shown.

In order to test the assumptions made by Kendall and Hulin, one should sample workers from at least two communities. Measures of satisfaction with the job should be obtained from these samples of workers as well as their verbalizations regarding cost of living, what is regarded as an adequate wage, other jobs in the community, etc. From such data one would expect (a) across communities there would be an effect on job satisfaction similar to that observed by Hulin (1966) and Kendall (1963), (b) across communities there should be an effect on the verbalizations which are supposedly tapping the intervening variables, and (c) within communities there should be an effect on job satisfaction due to the differing responses made by the individual workers to the same situations. That is, workers who verbalize that their community has a high cost of living, for example, should be less satisfied with wages than those who say that the cost of living is low.

The research described in this paper was designed to test some of the assumptions made by Kendall and Hulin and the generality of the relationships between community characteristics and job satisfaction.

METHOD

Research sites

The research was conducted in two "company" towns in British Columbia, Canada. The larger of the two communities (Town A) has a population of 10,000 people. Although the company no longer owns

the stores, service areas, and individual houses, it does own all of the surrounding land, and building lots must be purchased from the company. Building plans for individual houses must be approved by a committee of company executives. The commercial establishments of the town, while technically independent of the company, are in fact regulated by company policy. While most of the more distasteful accoutrements of the traditional company town are absent (Allen, 1966), this town is so strongly dominated by one company and has enough of the characteristics to be considered a company town.

Town A is relatively isolated from the remainder of British Columbia. For example, it is located at a distance requiring a 2-hr. plane ride (if the plane flies), a long two-day drive, or a 30-hr. boat trip from the nearest major city. The physical setting of Town A is spectacularly beautiful. The town's public facilities are generally excellent. For example the town has a 350 bed hospital which was built by the company and sold to the community for a nominal sum. All of the town's streets are paved and all of the houses are on a sewer system and town water supply. Ten doctors (eight general practitioners and two obstetric and gynecology specialists) practice in the community. In contrast, there are only two dentists. The physical facilities in the community's educational system are also outstandingly good. The schools are located throughout the community and none of them is crowded. The adequacy of the recreational facilities in the community varies depending on the age of the participant. There is the usual collection of curling sheets, hockey rinks, softball and baseball diamonds, and football fields in the community. Many of these were built by the company. However, for the very young children and the post-high-school inhabitants, organized recreational facilities are inadequate. Recreational opportunities are plentiful for the adults if they appreciate hunting, fishing, camping, and boating. However, beyond this, there is very little. Fancy restaurants and night clubs are totally lacking. There is one movie theatre which operates three nights a week. As a side comment, it should also be pointed out that crime, juvenile delinquency, unemployment, and air pollution are also absent and the community enjoys one of the highest standards of living in Canada.

The other of the two communities (Town B) has a population of approximately 200 people. It is located 50 mi. by water from the larger community. There are no roads in or out of this town. Transportation in and out is either by a twice-weekly boat or, in case of emergency, by one of the helicopters stationed there. The physical setting for the smaller community is equally spectacular. It is located in a narrow canyon at the confluence of two mountain streams. There are notable differences, however, between this community and the larger community. In Town B all the houses are owned by the company and rented to the inhabitants. This is not a great hardship since a three-bedroom house rents for approximately \$35 per month. There is only one store in town which is owned by the

company, but leased to a third party. The company controls the profits of this third-party leasor. There is only one eight-grade school in this smaller town which requires that older children be boarded in the larger community during the school year where they attend high school. The family of high school age children receives a boarding allowance. The town has a small dispensary instead of a hospital with one public nurse fulfilling the role of town medical agent. There are no dentists or dental facilities in this community. These two communities, then, while both being towns controlled by the same company do differ in significant ways.

Such research sites insure that the workers sampled will be responding to the same community characteristics and not to characteristics of a suburb or satellite community. Also, all the workers are employed by the same company. The physical environment has been controlled and most of the unwanted variation in the social and economic environment has been effectively removed. Workers' responses in such communities would provide a partial test of Kendall's and Hulin's assumptions.

Subjects

All of the salaried workers in both communities were requested to take part in this study. This included everybody from the first line supervision to the plant manager. The questionnaires were actually completed by 76% of the salaried workers in the communities. This included 442 workers from Town A and 28 workers from Town B. The questionnaires were given to the individual workers by work-group representatives who had met previously with the investigator. The individual workers were requested to complete the questionnaire on company time and send it through company mails to the investigator or mail it directly to the University of Illinois. Anonymity was guaranteed to all the workers in the sample. The investigator remained in the community and worked out of an office remote from the personnel department for a period of 10 days. Any workers who wished to talk to the investigator were urged to do so. A great number of the workers did. The characteristics of the two samples of workers are given in Table 1.

Variables

On the face sheet of the questionnaires, questions were asked regarding the respondent's age, the department in which he worked, sex, years of formal education, length of service with the company, length of service at present location, job title, whether the worker was on shift work or day work, and work location (community). The job title was used to obtain measures of the worker's job level and salary.

Measures of the workers' satisfaction with actual work done, pay, promotional opportunities, supervision, and co-workers were obtained by means of the Job Descriptive Index (JDI). The JDI is a cumulative point adjective check-list measure of job

TABLE 1
SAMPLE CHARACTERISTICS OF WORKERS
FROM TWO COMMUNITIES

| Characteristic | Town A (N = 442) | | Town B (N = 28) | |
|-----------------------------|---------------------|----------------|--------------------|----------------|
| | \bar{X} | S ² | \bar{X} | S ² |
| Age | 39.71 | 11.06 | 39.79 | 12.85 |
| Sex (1 = male, 0 = female) | .83 | .14 | .75 | .19 |
| Tenure with company | 8.94 | 5.71 | 6.19 | 5.28 |
| Tenure at location | 6.55 | 5.08 | 5.12 | 4.51 |
| Shift (0 = shift, 1 = days) | .84 | .14 | .93 | .06 |
| Yrs. education | 12.51 | 2.75 | 12.11 | 3.77 |

satisfaction which possesses adequate convergent and discriminant validity for individual analysis (Quinn & Kahn, 1967, p. 456; Smith, 1967; Vroom, 1964, p. 100). Three additional job-related satisfaction variables were assessed by means of the 5-point graphic rating scale described below. The three additional variables were satisfaction with management's response to complaints, satisfaction with training opportunities, and satisfaction with working conditions.

Two measures of overall satisfaction were also obtained. Each worker's satisfaction with his job in general (JIG) and life in general (LIG) was estimated by means of the General Motors Faces Scale (Kunin, 1955). In this particular study 7-point versions of the original 11-point scale were used. Each of these two general satisfaction scales consisted of three smiling faces, one neutral face, and three scowling faces. The workers were asked to indicate how they felt about their job in general or life in general, considering everything about their present situation, by checking the appropriate face.

Workers' responses to their communities were assessed by 5-point graphic rating scales. The scale points were labeled Very Dissatisfied, Somewhat Dissatisfied, Neither Satisfied nor Dissatisfied, Somewhat Satisfied, and Very Satisfied. The workers were asked to indicate their degree of satisfaction with each of the characteristics listed. They were asked to place a check mark on the line at the point which indicated their degree of satisfaction. They were told that the mark could be made at any point along the line. These scales were used to assess workers' satisfaction with medical facilities, school facilities, the weather, availability of living accommodations, availability of doctors, recreational facilities, teachers in the school system, availability of dentists, shopping facilities, cost of living, recreational facilities for children, cost of housing, location of the community in terms of its isolation and remote location, and the attractiveness of the community as a place to live. The same 5-point graphic rating scale was used to assess the workers' satisfaction with management's

responsiveness to complaints, satisfaction with training opportunities, and satisfaction with physical conditions at their place of work.

Finally, the importance each worker ascribed to these same community and job characteristics in determining the way he felt generally was assessed by means of similar 5-point graphic rating scales. The intervals of these 5-point scales were labeled Very Important, Important, Only Moderately Important, Unimportant, Very Unimportant. In the present study only the analysis of the workers' satisfaction scores will be reported.

Hypotheses

Examination of the assumptions made by Hulin and Kendall in their analyses of the effects of community characteristics on job satisfaction involves testing two interrelated sets of hypotheses. The first set of hypotheses is concerned with the differences in the workers' reactions to the two communities. The second set of hypotheses is related to the relationships between the workers' reaction to the communities and their job satisfaction within the two samples. Since the two communities involved in this research differ on known characteristics it is possible to formulate *a priori* hypotheses regarding differences in the workers' reactions to the characteristics of the communities. Many of these hypothesized differences are based on an enumeration of the physical characteristics of the communities. Such differences may be revealed, for example, by the usual census-type data. Other hypotheses are based on the author's reaction to the two communities and on conversations with the workers.

It would be expected that the workers in Town A would be more satisfied with the medical facilities, school facilities, availability of doctors, availability of dentists, and shopping facilities than would the workers in Town B. On the other hand, the workers in Town B should be more satisfied with the availability of housing, the cost of living, the cost of housing. No differences were hypothesized for the remaining variables.

Confirmation of these expected differences in the two mean vectors of responses would be evidence that the workers were responding in the expected manner to community characteristics. Further, it would indicate that the hypothesized intervening variables through which community characteristics have their effect on job satisfaction were being affected in the appropriate direction by differences in community characteristics.

The second set of hypotheses is concerned with the relationships between the workers' responses to their communities and their satisfactions with specific aspects of the job. The only hypotheses which can be made with any degree of certainty involve the relationship between the workers' satisfaction with pay and their response to the economic factors of the community. Therefore, we would hypothesize that there would be positive relationships between the workers' pay satisfaction and their satisfaction with the availability of housing, satisfaction with shopping

facilities, satisfaction with cost of living, and satisfaction with cost of housing. There will probably be other relationships between responses to aspects of communities and specific aspects of job satisfaction but no predictions can be made at this time. Confirmation of this set of hypotheses would be evidence supporting the mediating effects of satisfaction with community characteristics on job satisfaction.

Finally one can attempt to predict the two measures of overall satisfaction using the complete set of community satisfaction measures and job satisfaction measures as predictors. While such an analysis is not necessary for confirmation of the frame of reference hypothesis, it does provide evidence regarding the effects of community and job characteristics on general job and life satisfaction. Here again, a number of hypotheses can be made. It would be predicted that for the total sample of workers, satisfaction with specific aspects of the job would be more strongly related to general job satisfaction than to general life satisfaction. On the other hand, it is expected that satisfaction with community characteristics would be more closely related to general life satisfaction than to general job satisfaction and that the multiple correlations predicting general job satisfaction would be larger than the multiple correlations obtained when predicting general life satisfaction. While this represents a relatively good job of sampling characteristics of the job known to be related to overall job satisfaction, the sampling of community characteristics may not include many of the most important characteristics which determine general life satisfaction. Finally, differences should be expected between the male and female samples of workers in terms of the variables which contribute to variation in overall life and job satisfaction. Such differences would confirm the differing motivational characteristics of the two groups and would provide evidence on the nature of the motivational differences between males and females.

RESULTS

The data necessary to test the first hypothesis regarding differences in the workers' responses to the two communities are presented in Table 2 which gives the means for each of the 14 community satisfaction variables for the two groups of workers. The *t* ratios testing the significance of the difference between the two groups are also presented for those variables for which predictions were made. It is recognized that this analysis technique is not completely appropriate for these data. While the alpha level for any particular comparison will be accurate, *t* tests on correlated dependent variables will give an underestimate of the alpha level for the study as a whole. However, the other candidates for the test of the differences between the mean vectors

TABLE 2
SATISFACTION WITH COMMUNITY CHARACTERISTICS

| Satisfaction with | Town A (N = 442) | | Town B (N = 28) | | <i>t</i> | <i>p</i> |
|-------------------------------|---------------------|----------------|--------------------|----------------|----------|----------|
| | \bar{X} | S ² | \bar{X} | S ² | | |
| Medical facilities | 4.21 | .93 | 3.00 | 1.22 | 5.58 | <.01 |
| School facilities | 4.00 | .97 | 3.39 | 1.03 | 3.03 | <.01 |
| Weather | 2.60 | 1.02 | 3.07 | 1.21 | | |
| Availability of housing | 3.04 | 1.44 | 4.14 | 1.27 | 4.94 | <.01 |
| Availability of doctors | 4.19 | .97 | 2.54 | 1.20 | 7.64 | <.01 |
| Adult recreational facilities | 3.55 | 1.21 | 4.18 | 1.12 | | |
| Teachers | 3.58 | .99 | 3.68 | 1.12 | | |
| Availability of dentists | 2.71 | 1.30 | 1.89 | 1.03 | 4.04 | <.01 |
| Shopping facilities | 2.45 | 1.14 | 1.86 | .93 | 3.07 | <.01 |
| Cost of living | 2.14 | 1.00 | 3.39 | 1.34 | 5.49 | <.01 |
| Children's recreation | 3.65 | 1.07 | 3.25 | 1.17 | | |
| Cost of housing | 2.52 | 1.32 | 4.64 | .62 | 13.15 | <.01 |
| Location of community | 2.90 | 1.22 | 3.32 | 1.25 | | |
| Attractiveness of community | 3.97 | .75 | 4.29 | .76 | | |

(*T*² analysis, discriminant function analysis), while avoiding the issue of correlated dependent variables, are unjustified on other grounds (unequal variance-covariance matrices).

It can be seen that for the eight variables where differences were predicted, the differences did occur, were always in the expected direction, and were sizable enough to reach significance. Significance, however, may be of little practical value since the total *N* is 470. Such an *N* almost ensures statistical significance for even small differences. The size of the difference in comparison to the variance of the variables is usually large. For the six variables where no differences were predicted, the differences are generally smaller. Such results confirm the effects of community characteristics on responses to the communities. Further, regarding responses to the communities as intervening variables, such results are evidence that the differences between communities have the desired effect on the differences in the intervening variables.

It is also interesting to note that the differences in the responses to the two communities not only confirm the hypothesis but also that the relative magnitudes of the satisfaction measures within each of the communities are as expected. Those characteristics considered to be outstanding in Town A, such as medical

facilities, school facilities, and the availability of doctors, are in fact the three variables which obtained the highest mean satisfaction scores. Those variables considered to be worse, such as availability of dentists, cost of living, and cost of housing, did in fact receive very low satisfaction scores from these workers. The characteristics considered most outstanding in Town B were the availability of housing and the cost of housing. These two variables received two of the three highest satisfaction scores. The three characteristics in Town B considered outstandingly bad were the availability of doctors, availability of dentists, and shopping facilities. These variables also received the lowest mean satisfaction scores. Therefore the data in Table 2 not only confirm the differences between communities, but also the relative magnitudes of the satisfaction scores within communities are in line with expectation, further evidence for the lawful operation of these intervening variables.

Tables 3 and 4 present the correlations between the workers' responses to their communities and their job satisfactions. It was hypothesized that the workers' satisfaction with the availability of housing, shopping facilities, cost of living, and cost of housing would have positive relationships with their satisfaction with pay. These four community

TABLE 3

CORRELATIONS BETWEEN RESPONSES TO COMMUNITY AND SPECIFIC JOB SATISFACTION MEASURES IN TOWN A

| Satisfaction with | Satisfaction with | | | | |
|-------------------------------|-------------------|------------------|------------|------------|------------|
| | Work | Pay | Promotions | Supervisor | Co-workers |
| Medical facilities | .19 | .06 | .02 | .11 | .17 |
| School facilities | .15 | .10 | .04 | .12 | .17 |
| Weather | .17 | .14 | .06 | .08 | .21 |
| Availability of housing | .18 | .24 ^a | -.04 | .13 | .23 |
| Availability of doctors | .10 | .11 | -.01 | .09 | .20 |
| Adult recreational facilities | .21 | .23 | .04 | .18 | .26 |
| Teachers | .06 | .06 | .04 | .11 | .16 |
| Availability of dentists | .09 | .18 | .05 | .12 | .19 |
| Shopping facilities | .12 | .26 ^a | .02 | .18 | .16 |
| Cost of living | .19 | .38 ^a | .14 | .18 | .22 |
| Children's recreation | .24 | .20 | .10 | .20 | .20 |
| Cost of housing | .16 | .30 ^a | .02 | .16 | .20 |
| Location of community | .18 | .12 | .08 | .14 | .19 |
| Attractiveness of community | .12 | .16 | .05 | .14 | .18 |

Note.—For an *N* of 442, $r > .10$, $p < .05$; $r > .13$, $p < .01$.

^a Predicted to be positive and significant.

variables are reflecting (or are related to) the cost of living in the community. Previous results indicate that variables indexing cost of living have their strongest relationships with pay satisfaction. Variables measuring satisfaction with cost of living are expected to have strong relationships with pay satisfac-

tion. The data in Table 3, based on the sample of workers from Town A, support this hypothesis. Satisfaction with adult recreational facilities and children's recreational facilities are the other two variables having strong relationships with pay satisfaction. Why these two variables should be related to

TABLE 4

CORRELATIONS BETWEEN RESPONSES TO COMMUNITY AND SPECIFIC JOB SATISFACTION MEASURES IN TOWN B

| Satisfaction with | Satisfaction with | | | | |
|-------------------------------|-------------------|------------------|------------|------------|------------|
| | Work | Pay | Promotions | Supervisor | Co-workers |
| Medical facilities | -.08 | .16 | .09 | .28 | -.10 |
| School facilities | .27 | .40 | .25 | .36 | .21 |
| Weather | .22 | -.06 | -.03 | -.12 | .28 |
| Availability of housing | .06 | .25 ^a | .17 | -.05 | -.09 |
| Availability of doctors | -.12 | -.21 | -.04 | .04 | -.18 |
| Adult recreational facilities | -.10 | -.01 | -.17 | -.22 | -.02 |
| Teachers | .27 | .31 | -.15 | .33 | .28 |
| Availability of dentists | .23 | .00 | -.08 | .11 | .06 |
| Shopping facilities | .23 | .03 ^a | .21 | -.15 | .15 |
| Cost of living | .23 | .44 ^a | .05 | .14 | .11 |
| Children's recreation | -.12 | .27 | -.01 | .05 | -.07 |
| Cost of housing | -.09 | .55 ^a | .33 | .23 | -.29 |
| Location of community | .28 | .12 | .01 | .21 | .36 |
| Attractiveness of community | -.11 | .33 | .25 | .32 | .20 |

Note.—For an *N* of 28, $r > .37$, $p < .05$; $r > .47$, $p < .01$.

^a Predicted to be positive and significant.

TABLE 5

CORRELATIONS OF ALL SATISFACTION VARIABLES WITH
JOB IN GENERAL AND LIFE IN GENERAL OF MALES

| Satisfaction variable | JIG | LIG |
|-------------------------------------|-------|-------|
| JDI work | .43** | .27** |
| JDI pay | .25** | .20** |
| JDI promotion | .21** | .15** |
| JDI supervisor | .32** | .13* |
| JDI co-workers | .31** | .20** |
| Management's response to complaints | .32** | .22** |
| Training opportunities | .16** | .09 |
| Working conditions | .27** | .16** |
| Medical facilities | .04 | .09 |
| School facilities | .03 | .12* |
| Weather | .20** | .25** |
| Housing | .22** | .23** |
| Doctors | .07 | .13* |
| Adult recreational facilities | .15** | .22** |
| Teachers | .04 | .15** |
| Dentists | .14* | .13* |
| Shopping facilities | .19** | .22** |
| Cost of living | .20** | .23** |
| Child recreation | .19** | .12* |
| Cost of housing | .11 | .09 |
| Location of town | .27** | .30** |
| Attractiveness of town | .22** | .24** |

Note.— $N = 387$.* $p < .05$; two-tailed test.** $p < .01$; two-tailed test.

pay satisfaction is unclear. It should be noted, however, that these two responses to the community show the most general relationships with all of the measures of specific job satisfaction. The data in Table 3 also indicate moderate relationships between satisfaction with community characteristics and satisfaction with actual work done. This is in line with previous results. Not in line with previous results are the relationships between community characteristics and co-worker satisfaction. Satisfaction with co-workers generally has not behaved in the lawful manner that the other job satisfaction variables have. It is usually unrelated to the predictor variables which are useful in understanding variation in other aspects of job satisfaction.

The results in Table 4 based on the workers from Town B give only limited support to the second hypothesis. Three of four community variables hypothesized to have positive relationships with pay satisfaction tend to have high relationships but only two are significant. Satisfaction with shopping facilities does not fall in line with the remainder of

the results. In addition, satisfaction with school facilities and satisfaction with the attractiveness of the community also show substantial relationships with pay satisfaction. It should be pointed out that these results are based on a sample of only 28 workers. The reliability of the correlations is not high.

Taken together, the data presented in Tables 3 and 4 generally support the second hypothesis. The two samples were combined and all future results will be based on the combined sample of workers from the two towns.

Tables 5 and 6 present the relationships between all of the job satisfaction variables and the community satisfaction variables on the one hand and general job and life satisfaction on the other. Table 5 contains the data from the male workers and is based on an N of 387. The five JDI satisfaction variables and the three graphic rating scales assessing satisfaction with specific aspects of the job generally have positive relationships with general job satisfaction. Seven of these job satisfaction

TABLE 6

CORRELATIONS OF ALL SATISFACTION VARIABLES
WITH JOB IN GENERAL AND LIFE
IN GENERAL OF FEMALES

| Satisfaction variable | JIG | LIG |
|-------------------------------------|-------|-------|
| JDI work | .55** | .14 |
| JDI pay | .29** | -.01 |
| JDI promotion | .30** | -.01 |
| JDI supervisor | .39** | -.03 |
| JDI co-workers | .37** | .23* |
| Management's response to complaints | .45** | .06 |
| Training opportunities | .08 | .21 |
| Working conditions | .40** | .12 |
| Medical facilities | .01 | .05 |
| School facilities | .12 | .08 |
| Weather | .12 | .23* |
| Housing | .27* | .08 |
| Doctors | .09 | .03 |
| Adult recreational facilities | .10 | .20 |
| Teachers | .07 | .07 |
| Dentists | .03 | -.01 |
| Shopping facilities | .22* | .10 |
| Cost of living | .22* | .08 |
| Child recreation | .07 | .27* |
| Cost of housing | .16 | .15 |
| Location of town | -.02 | .29** |
| Attractiveness of town | .14 | .09 |

Note.— $N = 80$.* $p < .05$; two-tailed test.** $p < .01$; two-tailed test.

TABLE 7
VARIMAX ROTATED FACTOR MATRIX OF WORK-RELATED VARIABLES OF MALES

| Satisfaction variable | Factor | | |
|--|--------------------------------------|-----------------|------------------|
| | I Inter- personal relations | II Intrinsic | III Extrinsic |
| JDI work | .36 | .70 | .11 |
| JDI pay | .04 | .08 | .91 |
| JDI promotion | .03 | .84 | .10 |
| JDI supervision | .56 | .38 | .29 |
| JDI co-workers | .68 | .38 | .07 |
| Management's response to complaints | .65 | .06 | .50 |
| Training opportunities | .79 | .06 | .02 |
| Working conditions | .29 | .35 | .44 |
| % Total variance explained | 25 | 20 | 17 |
| % Common variance explained | 40 | 32 | 27 |

Note.—N = 387.

variables also have significant relationships with LIG for the male workers but the relationships are all lower. In addition, it can be seen that the community satisfaction variables tend to have stronger relationships with LIG satisfaction than they do with JIG satisfaction. There are 3 out of 14 reversals among these relationships, however, and the differences between the correlations are usually small.

The data in Table 6 are somewhat different. The relationships shown in Table 6 are based on the sample of 82 female workers. Seven of the eight job satisfaction variables again show strong relationships to JIG satisfaction. However, these seven variables are the only variables which bear substantial relationships to satisfaction with JIG for this sample. The only other variables which are significantly correlated with JIG for females are three of the four variables assumed to be reflecting satisfaction with the cost of living in the communities. It can also be seen that for female workers the predictability of LIG satisfaction is limited with only four variables having significant correlations with this variable. These significant correlations also tend to be rather low.

The interpretation of the relationships in Tables 3, 4, 5, and 6 is complicated by the presence of so many variables. The community- and job-related satisfaction variables were subjected to principal component analyses for the male and female workers separately, in order to simplify the interpretations. Table 7 presents the three-factor Varimax rotated solution for the eight work-related variables for the male sample. (In this and all subsequent factor analyses in this study, principal axis solutions with unities in the main diagonal were used. In all cases several different numbers of factors were rotated to the Varimax criterion. The solutions presented are the ones which appeared best in summarizing the data and making psychological sense.) These three factors explained 62% of the total variance for males. The first factor was interpreted either as a general factor or as an interpersonal relations satisfaction factor. JDI supervision satisfaction, JDI co-worker satisfaction, satisfaction with management's responsiveness to complaints, and satisfaction with training programs all involve some element of the workers' response either to their co-workers or to the company management in the plant. The second factor was interpreted as an intrinsic job satisfaction factor since it was defined by high loading of JDI work and JDI promotion (advancement) satisfaction. The third factor was interpreted as an extrinsic job satisfaction factor and was defined by high loadings for JDI pay satisfaction with working conditions in the plant.

The 14 community satisfaction variables for the male workers were also factor analyzed. These results are presented in Table 8. The five components were rotated to the Varimax criterion. These five factors explained 65% of the total variance. The first factor extracted was interpreted as an economic factor and was defined by high loadings of satisfaction with availability of housing, satisfaction with shopping facilities, satisfaction with the cost of living, and satisfaction with the cost of housing. The second factor was interpreted as satisfaction with the physical setting of the community. It was defined by high loadings of satisfaction with the weather in the community, satisfaction with

TABLE 8
VARIMAX ROTATED FACTOR MATRIX OF COMMUNITY VARIABLES OF MALES

| Satisfaction variable | Factors | | | | |
|---------------------------------|--------------------------|---------------------------|------------------------------|--------------------------------|------------------------------|
| | I Economic factors | II Physical setting | III Medical facilities | IV Recreation facilities | V Education facilities |
| Medical facilities | -.05 | .17 | .77 | .25 | .17 |
| School facilities | .10 | .04 | .36 | .14 | .74 |
| Weather | .17 | .81 | .11 | -.01 | .02 |
| Housing availability | .76 | .13 | .08 | .08 | .20 |
| Doctor availability | .06 | .12 | .84 | .07 | .04 |
| Adult recreational facilities | .28 | .16 | .08 | .68 | .14 |
| Teachers | .13 | .09 | -.03 | .12 | .86 |
| Dentist availability | .51 | -.09 | .52 | -.12 | .14 |
| Shopping facilities | .59 | .21 | .24 | .25 | .02 |
| Cost of living | .57 | .43 | -.12 | .19 | -.10 |
| Children's recreation | .02 | .09 | .12 | .85 | .11 |
| Cost of housing | .79 | .20 | -.08 | .05 | .09 |
| Location of community | .24 | .78 | .03 | .11 | .06 |
| Community attractiveness | .08 | .54 | .21 | .27 | .13 |
| % Total variance accounted for | 17 | 14 | 13 | 11 | 10 |
| % Common variance accounted for | 26 | 21 | 20 | 17 | 16 |

Note.— $N = 387$.

the attractiveness of the community, and satisfaction with the location of the community. The third factor was interpreted as satisfaction with medical facilities in the community and was defined by high loadings of satisfaction with availability of dentists. The fourth factor was labeled as satisfaction with recreational facilities in the community and was defined by high loadings of satisfaction with adult recreation, and satisfaction with children's recreational facilities. The fifth factor was interpreted as satisfaction with the educational facilities in the community factor. It was defined by high loadings of satisfaction with school facilities and the satisfaction with the teachers in the school system. Factor scores were computed for the male workers on these eight factors.² Multiple regression analyses were then done between these eight factor scores as predictor variables and JIG and LIG as two criterion

variables. The results of these analyses are shown in Table 9. These results indicate that the job satisfaction variates have substantial relationships (in terms of standard partial beta weights) with JIG satisfaction and lower relationships with LIG satisfaction. On the other hand, only one of the five community factors (economic factors in the community) has a sizable relationship with LIG satisfaction and none is strongly related to JIG satisfaction.

TABLE 9
STANDARDIZED PARTIAL REGRESSION WEIGHTS
PREDICTING JOB IN GENERAL AND LIFE IN
GENERAL SATISFACTION OF MALES

| Predictor variable | JIG | LIG |
|--------------------------------------|------|------|
| Interpersonal relations | .42 | .13 |
| Intrinsic job satisfaction | .33 | .17 |
| Extrinsic job satisfaction | .24 | .13 |
| Economic factors in communities | .15 | .25 |
| Physical setting of communities | .06 | .12 |
| Medical facilities in communities | -.15 | -.01 |
| Recreation facilities in communities | .01 | .08 |
| School facilities in communities | .01 | .12 |
| R | .55* | .44* |

Note.— $N = 387$.
* $p < .01$.

² The formula used for computing the two sets of factor scores was $Z R^{-1} F$ where Z is on $N \times M$ data matrix of M standard scores for each of N Ss, R^{-1} is the inverse of the $M \times M$ intercorrelation matrix, and F is the $M \times K$ factor matrix of the loadings of the M variables on the K orthogonal factors.

TABLE 10
VARIMAX ROTATED FACTOR MATRIX OF WORK
RELATED VARIABLES OF FEMALES

| Satisfaction variables | Factors | |
|-----------------------------------|--------------|----------------------|
| | I General | II Ex- trinsic |
| JDI work | .67 | .38 |
| JDI pay | .05 | .76 |
| JDI promotion | .67 | .21 |
| JDI supervisor | .18 | .73 |
| JDI co-workers | .66 | .32 |
| Management response to complaints | .54 | .52 |
| Training opportunities | .77 | -.09 |
| Working conditions | .24 | .61 |
| % Total variance accounted for | 29 | 26 |
| % Common variance accounted for | 53 | 47 |

Note.—N = 81.

faction. For this sample of 387 male workers the multiple correlation between the eight predictor variables and JIG satisfaction is .55. The multiple correlation predicting LIG satisfaction is .44. Both of these multiple correlations are significant, $p < .01$.

The intercorrelations of the eight job-related variables for the female workers were also factor analyzed. Two components from this analysis were rotated to the Varimax criterion. The Varimax rotated matrix is presented in Table 10. The first factor extracted appears to be basically a general factor with high loadings for JDI work, JDI promotions, JDI co-workers, management's responsiveness to complaints, and satisfaction with training opportunities. The second factor is clearly an extrinsic job satisfaction factor. It is defined by high loadings of JDI pay, JDI supervisor, and satisfaction with working conditions. These two factors explained 55% of the total variance.

The intercorrelations of the 14 community satisfaction variables were also factor analyzed. Again a five-factor solution appeared to be the best way of summarizing these data. The factor matrix is presented in Table 11. These five factors appear to be identical to the five factors extracted from the data provided by the male sample although they were extracted in a different order. The factors were interpreted as satisfaction with medical

TABLE 11
VARIMAX ROTATED FACTOR MATRIX OF COMMUNITY VARIABLES OF FEMALES

| Satisfaction variable | Factors | | | | |
|---------------------------------|----------------------------|---------------------------|----------------------------|--------------------------------|------------------------------|
| | I Medical facilities | II Economic factors | III Physical setting | IV Recreation facilities | V Education facilities |
| Medical facilities | .90 | .02 | .09 | .06 | .07 |
| School facilities | .53 | .06 | -.14 | .13 | .60 |
| Weather | -.10 | .06 | .79 | -.00 | .11 |
| Housing availability | .06 | .87 | .02 | .08 | .18 |
| Doctor availability | .88 | .06 | -.04 | .08 | .08 |
| Adult recreational facilities | .04 | .29 | .05 | .81 | -.15 |
| Teachers | .03 | .10 | .12 | -.04 | .90 |
| Dentist availability | .59 | .09 | .25 | -.50 | -.06 |
| Shopping facilities | .50 | .34 | .34 | -.04 | .12 |
| Cost of living | .08 | .78 | .31 | .09 | -.06 |
| Children recreation | .14 | -.06 | .38 | .66 | .33 |
| Cost of housing | -.05 | .89 | .11 | .05 | .02 |
| Location of community | .13 | .18 | .63 | .00 | .03 |
| Community attractiveness | .16 | .14 | .72 | .22 | -.10 |
| % Total variance accounted for | 18 | 17 | 15 | 10 | 10 |
| % Common variance accounted for | 26 | 25 | 21 | 15 | 14 |

Note.—N = 81.

facilities, satisfaction with the economic factors in the community, satisfaction with the physical setting of the community, satisfaction with the recreational facilities in the community, and satisfaction with educational facilities in the community. These five factors explain 70% of the total variance of the original 14 × 14 intercorrelation matrix.

Scores were computed on each of these seven (two work-related and five community-related) factors for each of the 81 female workers. Two multiple regression analyses were carried out to examine the relationship between the seven predictor variables and JIG and LIG satisfaction. The results of these analyses are given in Table 12. An examination of Table 12 indicates that of the seven predictor variables only two bear substantial and significant relationships to JIG. These are intrinsic job satisfaction and extrinsic job satisfaction. The multiple correlation of .67 for predicting JIG satisfaction is highly significant. The results for predicting LIG satisfaction are neither consistent nor impressive. None of the beta weights for the predictor variables is significant and the multiple correlation, .36, was not significant.

Canonical regression analyses were also carried out predicting JIG and LIG job satisfaction jointly for male and female workers. These two canonical analyses were performed in order to examine the possible interactive effects of JIG and LIG satisfaction. The results of these two analyses in no way made the interpretation of the data any more clear or consistent. They will not be reported here.

DISCUSSION

The results of this study lend substantial support to the validity of the assumptions made by Kendall (1963) and Hulin (1966) in their analysis of the effects of community characteristics on job satisfaction. The data in this study demonstrate that differences between communities result in predictable differences in the workers' satisfaction with these communities. These same community characteristics which result in differences in their satisfaction with the cost of living in the community also have a significant effect on their satisfaction with pay. Secondly, it

TABLE 12
STANDARDIZED PARTIAL REGRESSION WEIGHTS
FOR PREDICTING JOB IN GENERAL AND LIFE
IN GENERAL SATISFACTION OF FEMALES

| Predictor variable | JIG | LIG |
|----------------------------|------|------|
| Intrinsic job satisfaction | .50 | .20 |
| Extrinsic job satisfaction | .55 | .07 |
| Medical facilities | -.21 | -.06 |
| Economic factors | -.05 | -.02 |
| Setting | -.02 | .21 |
| Recreational facilities | -.01 | .17 |
| Educational facilities | .02 | -.01 |
| R | .67* | .36 |

Note.—N = 81.
* p < .01.

was demonstrated that the workers' satisfaction with the economic characteristics of the community had the expected effect on their satisfaction with pay. Thirdly, it was demonstrated that workers' satisfaction with community characteristics and satisfaction with job characteristics considered jointly had significant and predicted effects on their satisfaction with their JIG and their satisfaction with their LIG. Finally differences between male and female workers in terms of the variables which were related to overall job and life satisfaction were reasonable. The magnitude of multiple correlations predicting general job satisfaction from community and specific job satisfaction variables were not only significant but were substantially large. For males, the eight predictor variables controlled approximately 30% of the variance in job satisfaction and for the females, the seven predictor variables accounted for approximately 45% of the variance. Therefore, not only are the relationships statistically significant, but they are large enough to be considered as practically significant.

Kendall (1963), Hulin (1966), and Smith, Kendall, and Hulin (1969) have maintained that the standards against which workers compare the level of their environmental return cannot be considered to be constant from one community to another, from one plant to another within the same community, or even from one worker to another within the same plant. These writers have stressed that using an adaptation level or frame of

reference set of hypotheses to analyze workers' responses to their jobs gives a much more consistent and meaningful set of results than simply assuming that what we as middle-class white investigators regard as good is what all workers will regard as good. The results of a small number of studies now done on this problem indicate that the assumptions made by Kendall, Hulin, and Smith are valid and are useful in understanding workers' responses to their jobs. Present data along with these previous studies also indicate that situational variables should no longer be considered as nuisance parameters to be controlled or partialled out of our predictive equations. They should be regarded as a valid and meaningful source of variance and their effects should be analyzed rather than removed.

Finally, these results in conjunction with the results of Blood and Hulin (1967) and Hulin and Blood (1968) indicate that not only must the effects of differences in preference for work role outcomes be considered but also differences in standards for judging the goodness or badness of any given level of any given work role outcome. While such a consideration undoubtedly complicates the life of the investigator, the benefits to be gained from such analyses are enormous. The ubiquitous sex differences in both job satisfaction and job motivation which have plagued investigators for years can probably be considered as simply a combination of differences in preferences for work role outcomes and standards for judging these work role outcomes. Likewise the earlier results of Katzell, Barrett, and Parker (1961), Turner and Lawrence (1965), and Whyte (1955) reporting urban-rural differences in job motivation are understandable as a combination of both differences in preference and difference in standards. Further, by using one set of moderator variables chosen to index individual differences in preferences for work role outcomes and another set of moderator variables designed to index internal standards for the judgment of levels of work role outcomes, data from large groups of industrial workers employed by different companies and in different communities can be analyzed without

resorting to laborious subgroup analyses. If such moderator variables are wisely chosen and accurately measured, moderated regression analyses predicting either motivation or job satisfaction should be useful for understanding the motivation of industrial workers.

Finally, the use of the implications contained in these data in conjunction with instrumentality theory (Vroom, 1964) may provide a means of linking these two approaches to the study of job satisfaction. The instrumentality theory as discussed by Vroom is completely ahistorical. No concern is paid to the sources of variance in job satisfaction, why different workers have different valences, or why two workers will attribute different instrumentalities to the same job for providing some second-level outcome. Instrumentality theory does provide a link between attitudes and behaviors. The traditional model, on the other hand, is basically concerned with the development of high or low job satisfaction. It provides no link between attitudes and behaviors. However, these two approaches could be easily and usefully combined. The traditional model could be used to predict the valence and instrumentality of Variable i for homogeneous subsets of workers. These predicted valences and instrumentalities could then be combined to obtain $\sum V_i I_i$ (a measure of overall job satisfaction). Such an estimate of overall satisfaction could be validated. The advantage of such an approach would be that all the variables would be experimentally independent of S . Therefore, use of such a combination would break the sterile confines of the ahistorical instrumentality model and at the same time avoid the problems of response-response laws.

REFERENCES

- ALLEN, J. B. *The company town in the American West*. Norman: University of Oklahoma Press, 1966.
- BLOOD, M. R., & HULIN, C. L. Alienation, environmental characteristics, and worker responses. *Journal of Applied Psychology*, 1967, **51**, 284-290.
- HULIN, C. L. The effects of community characteristics on measures of job satisfaction. *Journal of Applied Psychology*, 1966, **50**, 185-192.
- HULIN, C. L., & BLOOD, M. R. Job enlargement, individual differences, and worker responses. *Psychological Bulletin*, 1968, **69**, 41-55.

- KATZELL, R. A., BARRETT, R. S., & PARKER, T. C. Job satisfaction, job performance, and situational characteristics. *Journal of Applied Psychology*, 1961, **45**, 65-72.
- KENDALL, L. M. Canonical analysis of job satisfaction and behavioral, personal background, and situational data. Unpublished doctoral dissertation, Cornell University, 1963.
- KUNIN, T. The development of a new type of attitude measure. *Personnel Psychology*, 1955, **8**, 65-78.
- QUINN, R. P., & KAHN, P. L. Organizational psychology. *Annual Review of Psychology*, 1967, **18**, 437-466.
- SMITH, P. C. The development of a method of measuring job satisfaction: The Cornell studies. In E. A. Fleishman (Ed.), *Studies in personnel and industrial psychology*. Homewood, Ill.: Dorsey, 1969.
- SMITH, P. C., KENDALL, L. M., & HULIN, C. L. *The measurement of satisfaction in work and retirement*. Chicago: Rand McNally, 1969.
- TURNER, A. N., & LAWRENCE, P. R. *Industrial jobs and the worker: An investigation of response to task attributes*. Boston: Harvard University, 1965.
- VROOM, V. *Work and motivation*. New York: Wiley, 1964.
- WHYTE, W. F. *Money and motivation: An analysis of incentives in industry*. New York: Harper, 1955.

(Received May 31, 1968)

VOCATIONAL CHOICE AND PROFESSIONAL EXPERIENCE AS FACTORS IN OCCUPATIONAL IMAGE

EDMOND MARKS AND SAM C. WEBB¹

Georgia Institute of Technology

Trait descriptions of the typical incumbent of two distinct occupational titles were obtained from three groups varying in professional experience: (a) freshman students beginning their occupationally relevant education, (b) senior students completing this education, and (c) postgraduates working in the occupation. In addition, self-ratings on the same traits were obtained from a sample of freshman students enrolled in the two majors. The self- and other-ratings were also related to the social desirability (SD) ratings of the set of trait descriptions. The results indicated that the three groups varying in professional experience share a common "image" of the typical occupational incumbent, with this image being substantially related to the self-characterizations of freshmen enrolled in that major. Although SD ratings were highly related to the average trait characterizations, it was suggested that not all the differences in these latter responses could be accounted for in terms of socially desirable response tendencies.

Many theories of vocational choice, adjustment, and development focus upon the notion that vocational behavior—preference, choice, and performance—can be explained in terms of the congruence between the pattern of attributes or traits of the individual exhibiting the behavior and the pattern of attributes or traits of some external model, for example, the typical incumbent of an occupational category like lawyer, engineer, etc., or students of a given achievement level (Bordin, 1943; O'Dowd & Beardslee, 1960; Super, 1963; Tiedeman & O'Hara, 1963).

In this study, the effects of two factors on the description of the personal characteristics of the typical incumbent of a representative occupation made by Ss listing that occupation as their vocational choice were examined. The two factors studied were occupational title—for example, electrical engineering or architecture—and amount of training or experience related to vocational preparation. In addition to these descriptions of some idealized or typical "other" made by the groups studied, the self-characterizations held by a sample of entering college freshmen who listed the occupation title as their college major were obtained. Of particular interest was the congruence between these self-descriptions and

those of the typical incumbent of the occupational title.

Previous research (O'Dowd & Beardslee, 1960) suggests that college students possess fairly reliable stereotypes of a wide variety of high-level occupations and that these stereotypes differ substantially among various occupations. Changes in occupational images related to time or experience, on the other hand, have not been demonstrated. O'Dowd and Beardslee (1960), who compared freshman and senior images of selected high-level occupations, found little difference between these two levels of educational development. The failure to obtain a reliable relationship between experience and change in occupational image may well be due to the interval studied. The greatest and perhaps most rapid change may emerge after the individual enters the world of work. Also of interest in this study was the role of social desirability in self- and other-description.

METHOD

Procedure

Eight groups of Ss were established based partly on the cross-classification of the occupational title and professional experience factors. Actually, two of the groups fell outside the 2 × 3 cross-classificatory scheme. The two occupational titles selected for study were Industrial Management and Electrical Engineering. Rather specific titles were used in an effort to avoid confounding of possible differences

¹ Requests for reprints should be sent to Sam C. Webb, Dean, Division of Graduate Studies, Georgia Institute of Technology, Atlanta, Georgia 30332.

among occupational incumbents all falling within a larger title, engineering, for example. Although a larger number of occupational titles were actually sampled, these two were selected, on the basis of representativeness, to simplify the analysis.

The three levels of amount of training or experience established were freshman, senior, and professional. The freshman level consisted of students who had not begun any formal college course work, but who listed either industrial management or electrical engineering as their choice of major. Seniors were those students in these two occupational titles who were scheduled to graduate within either of these majors at the next commencement. The "professionals" were Ss who had obtained at least a baccalaureate degree in one of the two occupational titles and who were presently working in that occupation.

The two final groups were freshmen—again not having undertaken any formal college course work—who listed either of the occupational titles as their choice of major. These Ss were asked to describe themselves in terms of the same trait-descriptions used by other groups.

The six groups comprising the 2×3 classification defined by occupational title and professional experience were administered a 95-item trait-description instrument. The 95 items were selected to cover a broad range of behaviors or characteristics that a person might exhibit and which would be relevant to occupational endeavors. Some sample items are high intellectual ability, self-assurance, flexibility of thought, social immaturity, good control of impulses, and proneness to anxiety. The S was asked to rate each characteristic according to its importance in determining the success or failure of an individual working in the occupation that S had selected as his vocational choice. The S indicated his item response on a 9-point rating scale ranging from (1) "a factor of very great importance in determining failure," through (5) "a factor that would have no bearing on an individual's success or failure," to (9) "a factor of very great importance in determining success."

The two groups of freshman Ss who responded with self-descriptions were asked to rate the same 95 items with respect to how characteristic the given trait was of them. The Ss were asked to rate each item on a 9-point scale ranging from "very definitely not characteristic of me" through "neither characteristic nor not characteristic of me," to "very definitely characteristic of me."

Analyses

Several related analyses of the data were undertaken. In the first analysis the scale values and discriminant dispersions of each of the 95 trait terms were obtained using Thurstone's model for categorical judgment (Torgerson, 1958). The scale values and standard deviations were obtained separately for each of the eight groups by means of a graphical procedure for successive categories described by Bock and Jones (1963). Based upon these

scale values the trait descriptions falling in the two most extreme categories on both ends of the scale—Categories 1, 2, and 8, 9—were selected for further treatment. Items meeting this criterion could be considered as the most salient in characterizing the typical incumbent of the given occupation. To explore the notion of a common occupational stereotype among the four groups representing either of the occupational titles, the percentages of item overlap, and rank order correlations—ranked on scale value over the common items—were computed for the $n(n-1)/2$ pairs of groups within an occupational title. High values on both indexes would suggest that the groups within a given occupational title share a common image of the typical incumbent of that occupation. To provide an overall test of the hypothesis of a common occupational profile, a test described by Marks (1968) based on the principal components of the individual correlation matrices was employed.

Social Desirability

Another aspect of the present study considered relevant to the hypothesis of a common occupational image concerned the role of SD in responses to the trait items. The 95 trait names had been scaled by the method of successive categories from the responses from an independent sample of freshmen students. The rank order correlations between these SD scale values and the scale values obtained on each of the four groups within an occupational title were calculated. Should high correlational estimates obtain, Ss' "true" characterizations of an occupational incumbent might be confounded with their implicit notions of what constitutes socially desirable traits in an idealized person.

Subjects

A total of 674 Ss were used in the study. The breakdown was as follows. For industrial management there were 50 freshmen who engaged in self-description, and 65 freshmen, 84 seniors, and 81 professionals who rated a typical occupational incumbent. For electrical engineering there were 50 freshmen who engaged in self-description and 58 freshmen, 61 seniors, and 75 professionals who described the incumbent. In addition, a sample of 150 freshmen from another university rated the 95-trait descriptions for SD.

RESULTS

Level of Professional Experience

The results will be discussed first in terms of the professional experience dimension for the two occupational titles separately.

Industrial management. For the three levels of professional experience—freshman, senior, and professional—there were 43, 47, and 46 items, respectively, which fell into the two most extreme categories on the "importance"

TABLE 1
SCALE VALUES AND DISCRIMINAL DISPERSIONS^a OF THE 35 COMMON ITEMS FOR THE
INDUSTRIAL MANAGEMENT GROUPS

| Item | Freshman | Senior | Professional |
|---|------------|------------|--------------|
| 1. Excellent educational training | 1.83(.68) | 1.67(.60) | 1.75(.65) |
| 2. High level of physical energy | 1.11(.74) | 1.45(.70) | 1.65(.99) |
| 4. Efficient working habits | 1.46(.81) | 1.25(.75) | 1.75(.55) |
| 6. High intellectual ability | 1.83(.71) | 1.25(.75) | 1.60(.81) |
| 10. Inability to exercise authority | -.99(.99) | -1.23(.77) | -3.00(.99) |
| 14. Little confidence in ability | -1.00(.35) | -1.65(.99) | -3.00(.99) |
| 15. Conscientiousness | 1.08(.99) | 1.67(.92) | 1.80(.99) |
| 17. Self-assurance | 1.17(.89) | 1.67(.86) | 1.67(.60) |
| 18. Self-discipline | 1.17(.65) | 1.75(.67) | 1.88(.61) |
| 19. Inability to work well with others | -1.10(.65) | -1.65(.85) | -1.76(.99) |
| 27. Tactlessness | -1.26(.72) | -1.46(.98) | -1.69(.82) |
| 32. Flexibility of thought | 1.16(.69) | 1.92(.70) | 1.76(.67) |
| 33. Unawareness of the limiting factors in a situation | -1.83(.30) | -1.75(.55) | -1.17(.98) |
| 35. Inadequate verbal skills | -.84(.37) | -1.10(.75) | -1.80(.88) |
| 39. Poor deductive reasoning | -.84(.31) | -1.18(.57) | -1.81(.55) |
| 40. Good inductive reasoning | 1.11(.84) | 1.30(.68) | 1.48(.79) |
| 41. Aspiring to high levels of professional achievement | 1.50(.77) | 1.50(.80) | 1.50(.68) |
| 43. Ability to innovate | 1.07(.78) | 1.71(.73) | 1.85(.88) |
| 44. Ability to make appropriate judgments | 1.78(.79) | 2.18(.66) | 2.50(.90) |
| 45. Adaptability | 1.12(.69) | 1.98(.68) | 1.80(.63) |
| 46. Inability to gain confidence of others | -1.08(.98) | -1.73(.99) | -3.00(.97) |
| 49. Open-minded | 1.33(.68) | 1.65(.65) | 1.37(.95) |
| 55. Lack of persuasiveness | -.79(.40) | -1.25(.58) | -1.48(.90) |
| 58. Lack of foresight | -.93(.30) | -1.13(.70) | -1.39(.94) |
| 60. Ingenuity | 1.45(.80) | 1.46(.66) | 1.88(.70) |
| 61. Unreliable | 1.25(.40) | 2.24(.66) | 2.11(.81) |
| 64. Emotional stability | 1.30(.80) | 1.50(.35) | 1.74(.85) |
| 68. Perseverance | 1.15(.71) | 1.50(.90) | 1.85(.68) |
| 79. Inability to complete assignments and consistently meet deadlines | -1.23(.44) | -1.88(.99) | -2.50(.99) |
| 80. Consistent performance at top ability | 2.23(.57) | 2.01(.75) | 2.23(.69) |
| 81. Lack of insight into behavior of others | -.86(.50) | -1.12(.63) | -2.02(.99) |
| 84. Good professional skills | 1.47(.47) | 1.68(.65) | 1.69(.67) |
| 85. Industrious | 1.70(.81) | 1.75(.57) | 1.90(.74) |
| 91. Aggressiveness | 1.36(.71) | 1.52(.89) | 1.73(.99) |
| 92. Accepts authority | 1.48(.85) | 1.42(.88) | 1.37(.78) |

^a Discriminal values in parentheses.

dimension; that is, they were considered descriptive of the typical occupational incumbent. Of these total numbers of items, 35 items were common to all three groups. That is, 81% of the freshman items, 74% of the senior items, and 76% of the professional items were shared with the other two groups. This suggests considerable overlap or common perceptions of the typical industrial management incumbent by the three groups. Of the items that did not match across all three groups, 63% of the freshman items, 83% of the senior items, and 63% of the professional

items were shared with at least one of the other groups. Eighty percent of the freshman single-matched items—items that were shared with one of the other two groups—were shared with the seniors, while 20% were shared with the professionals. Of the senior single-matched items, 40% were shared with the freshmen and 60% were shared with the professionals. Finally, of the professional single-matched items, 14% were shared with the freshmen and 86% were shared with the seniors. There appears to be a slight shift in terms of common stereotype as one moves from freshman

to senior to professional. The freshmen and professionals are most distinct in terms of occupational image, while the seniors take up a position somewhere between the two. These comments are limited simply to the total number and percentage of attributes or traits that the three groups agree are descriptive of the typical industrial management incumbent.

Another way of examining the notion of a common occupational image is through the degree of correspondence among the scale values for the 35 common items. A high degree of correspondence would indicate that the ratings of importance assigned to each trait description are quite similar for the three groups. The scale values and discriminial dispersions of the 35 items for the three groups are presented in Table 1. First, the pairwise rank order correlations among the 35 scale values were computed. The Spearman rho value for the freshman and senior arrays was .69, while the correlation estimate for the freshman and professional groups was .69. The correlation between scale values for the senior and professional groups was .87. These values complement the results of the item overlap analysis. Taken together, these results indicate substantial correspondence among the three levels of professional experience with respect to occupational image or stereotype, although the question of a shift in image with experience requires an overall test which is now described.

The results of the overall test (Marks, 1968) bore out the conclusions of a common occupational image for the three groups. To apply this test, the dispersion matrix of the 35 standardized variables was decomposed into between-group and within-group dispersion matrices for the total sample—that is, freshman, senior, and professional industrial management groups. A principal component analysis was applied to the total sample dispersion matrix and the principal components extracted. The within-group dispersion matrix of the principal variables was computed using CS_wC' , where C = the matrix of characteristic vectors and S_w is the within group dispersion matrix of standardized variables. CS_wC' was then diagonalized using the Gram-Schmidt process. The stepwise statistics $P_i = d_{iiw}/d_{iit}$ —where d_{iit} is the i^{th} character-

TABLE 2
STEPWISE CRITERIA AND THEIR SIGNIFICANCE LEVELS

| P_i | Parameters | Significance levels |
|-------|------------|---------------------|
| .984 | 227/2, 1 | .15 |
| .995 | 226/2, 1 | .56 |
| .989 | 225/2, 1 | .20 |
| .997 | 224/2, 1 | .71 |
| .996 | 223/2, 1 | .63 |

istic root of the total dispersion matrix and d_{iit} is the i^{th} diagonal element of the diagonalized CS_wC' —were computed for the first five components. The remaining roots were quite small. The test statistics and their corresponding beta distribution parameters and significance levels are presented in Table 2.

None of the test statistics reached the .05 level of significance which indicates that the three groups can be considered to attribute a common profile with respect to shape to the personal characteristics of the occupational incumbent.

Electrical engineering. A similar analysis was undertaken for the three groups rating the typical electrical engineer. For the three levels of professional experience there were 31, 33, and 36 items, respectively, which fell into the two most extreme categories on either end of the "importance" dimension. Of these total numbers of items, 29 items were common to all three groups. That is, 94% of the freshman items, 88% of the senior items, and 81% of the professional items were shared with the other two groups. As with the industrial management groups, this suggests considerable overlap of response or a common description of the electrical engineer. Of the very few items that did not match across all three groups, none of the freshman items, 75% of the senior items, and 43% of the professional items were shared with at least one of the other groups. As is obvious, only the senior and professional groups shared any single-matched items. On the basis of item overlap there appears to be a strikingly high degree of correspondence among the three levels of electrical engineering Ss with respect to characterization of the typical incumbent of this occupation.

TABLE 3
SCALE VALUES AND DISCRIMINAL DISPERSIONS^a OF THE 29 COMMON ITEMS FOR THE
ELECTRICAL ENGINEERING GROUPS

| Item | Freshman | Senior | Professional |
|---|------------|------------|--------------|
| 1. Excellent educational training | 2.65(.60) | 2.74(.38) | 2.80(.42) |
| 4. Efficient working habits | 1.70(.81) | 2.11(.61) | 2.43(.60) |
| 6. High intellectual ability | 1.60(.62) | 2.30(.31) | 3.00(.81) |
| 15. Conscientiousness | 1.37(.50) | 1.63(.62) | 2.00(.61) |
| 18. Self discipline | 1.36(.46) | 1.52(.61) | 1.80(.60) |
| 26. Inquisitiveness | 1.40(.51) | 2.06(.61) | 2.40(.57) |
| 32. Flexibility of thought | 1.63(.72) | 1.71(.81) | 2.07(.80) |
| 34. Good mathematical skills | 2.07(.64) | 2.31(.61) | 2.64(.37) |
| 37. Good mechanical comprehension | 1.43(.47) | 1.59(.61) | 1.72(.55) |
| 38. Analytical skills | 1.68(.43) | 2.50(.46) | 2.61(.63) |
| 39. Poor deductive reasoning | -1.10(.42) | -2.01(.48) | -2.17(.50) |
| 40. Good inductive reasoning | 1.42(.50) | 1.61(.81) | 1.96(.80) |
| 41. Aspiring to high levels of professional achievement | 1.45(.80) | 1.42(.57) | 1.40(.58) |
| 44. Ability to make appropriate judgments | 1.41(.79) | 1.46(.62) | 1.61(.60) |
| 45. Adaptability | 1.55(.90) | 1.39(.61) | 1.46(.60) |
| 49. Open-minded | 1.23(.81) | 1.36(.72) | 1.56(.70) |
| 58. Lack of foresight | -1.00(.61) | -1.01(.68) | -1.08(.63) |
| 60. Ingenuity | 1.73(.45) | 2.15(.50) | 2.91(.48) |
| 61. Unreliable | -1.38(.36) | -2.36(.40) | -2.39(.50) |
| 62. Exacting | 1.63(.28) | 1.91(.62) | 2.16(.48) |
| 68. Perseverance | 1.35(.46) | 1.50(.81) | 1.68(.62) |
| 75. Inability to isolate relevant feature within a body of data | -1.01(.69) | -1.09(.68) | -2.01(.63) |
| 76. Resistant to change | -1.07(.80) | -1.19(.58) | -1.22(.80) |
| 79. Inability to complete assignments and consistently meet deadlines | -1.32(.39) | -1.62(.68) | -1.52(.61) |
| 80. Consistent performance at top ability | 2.21(.32) | 1.80(.62) | 1.91(.70) |
| 83. Well read within profession | 1.36(.68) | 1.84(.60) | 2.31(.69) |
| 84. Good professional skills | 1.77(.61) | 1.98(.54) | 2.53(.67) |
| 85. Industrious | 1.65(.55) | 1.70(.57) | 1.86(.55) |
| 93. Superior high school and college achievement | 1.70(.41) | 2.01(.51) | 2.24(.57) |

^a Discriminal dispersions in parentheses.

This is further borne out by the pairwise correlations between scale values on the 29 common items. The scale values and discriminial dispersions of these 29 items for the three groups are presented in Table 3.

The correlation estimate between freshman and senior scale values was .84, while the correlation between freshman and professional

scale values was .79. The correlation between senior and professional values was .98. Again, these values complement the item overlap analysis by showing considerable correspondence between the rating of trait descriptions for the groups taken in pairs. The overall test, as described in the industrial management results, was applied to the three electrical engineering groups. The stepwise test statistics, and their beta distribution parameters and significance values are presented in Table 4. Only three characteristic roots were large enough to be considered.

As with the industrial management groups, there is no basis for rejecting the hypothesis of a common profile for the three electrical engineering groups. Taken together with the

TABLE 4
STEPWISE CRITERIA AND THEIR SIGNIFICANCE LEVELS

| P _i | Parameter | Significance levels |
|----------------|-----------|---------------------|
| .984 | 191/2, 1 | .20 |
| .991 | 190/2, 1 | .42 |
| .997 | 189/2, 1 | .74 |

item overlap and correlation data, these results indicate a similar occupational image for the groups varying in professional experience.

Self-description. Since the tests on the hypotheses relating to professional experience supported the notion of a common occupational image, the data for the three groups within an occupational title were pooled and the scale values of the 95 items for the two freshman self-descriptions were correlated with the scale values for the two pooled groups.

The Spearman rho between self-descriptions of the industrial management freshmen and the scale values for the pooled ratings of the typical occupational incumbent by the three industrial management groups varying in professional experience was .69. The corresponding value for the electrical engineering groups was .72. These values may not be comparable to the rho values reported previously for the two occupational titles. The latter values were based only on items common to all three groups—that is, items falling on the extremes of the “importance” dimension. Some attenuation of correlation estimates using such extreme categories would be expected. Nonetheless, the present rho values indicate substantial correspondence between the way a freshman student “sees” or describes himself and the common image ascribed by Ss at varying levels of professional experience to the typical occupational incumbent.

For the most part, the discriminial dispersions of the freshman self-descriptions tended to be larger than the same values for the three “experience” groups. The average absolute difference between these discriminial dispersions over the 95 items was .12. Only for 6 items was the freshman self-description discriminial dispersion less than the corresponding value for the “experience” groups. As such, self-descriptions demonstrated greater variability than other-descriptions.

Occupational Title

Since there were no significant differences among the three groups within an occupational title varying in professional experience, the data within an occupational title were

pooled and comparisons between the two titles were made on these common images.

In terms of item-overlap, of the 35 industrial management and 29 electrical engineering items falling into the previously defined extreme categories of the importance dimension, 20 items were common to both groups. That is, 57% of the trait-descriptions considered salient by the industrial management Ss and 69% of such items for the electrical engineers were shared by the groups. This is somewhat, but not much, less than the percentage overlap reported for the professional experience factor.

Although there is a moderate amount of overlap in the traits used to describe the typical occupational incumbent by the two groups, the pattern of judged importance of these common traits is quite different for the groups. The Spearman rho between scale values over the 20 common items was only .20. For all 95 items the same index was still only .32. Both values are quite low and suggest that the vocational images of the two titles are noticeably distinct.

In examining the trait descriptions that did not match, it was noted that the industrial management occupational image was more sensitive to and included trait descriptions involving interpersonal skills and traits, for example, inability to work well with others, emotional stability, and self-assurance. It appeared that the industrial management image was more complex in that it included the personality and interpersonal domains as well as the intellectual, training, and performance domains found in the electrical engineering image.

Social Desirability

The last aspect of the study investigated the relationship of the ratings assigned to the trait descriptions for the two occupational titles to the independent judgments of the social desirability of these traits. Several methods for examining this relationship have been proposed (Edwards, 1957), and the problems in interpretation of such correlation estimates have been discussed by Norman (1967). In this study, the item SD scale values (based upon judgments of 150 Ss)

were correlated over the 95 items with the item scale values of both the industrial management and electrical engineering groups separately. Most of the criticism leveled against this index by Norman (1967) is avoided by interpreting these correlations as reflecting properties of the items only, and not as involving *S* variability.

The Spearman rho's were .63 for industrial management and .43 for electrical engineering. Thus, a substantial relationship obtains between SD and "person" ratings for the industrial management title, and to a lesser extent for the electrical engineering group.

In terms of self-description, the rho values for the SD and freshman self-rating arrays—both industrial management and electrical engineering—were .64 and .61, respectively. These values were not noticeably different from those involving "other-description." Whatever the role of SD and self- and other-description, it appears to emerge equally in both types of tasks. Some possibilities relating to this role are discussed later.

DISCUSSION

The results obtained here are similar to those obtained from previous studies where different and broader occupational titles were used and shorter time intervals, and perhaps less qualitatively different levels of experience, were examined (O'Dowd & Beardslee, 1960). For one thing, amount of professional experience appears to have little effect upon an individual's image of the incumbent of an occupation for which the individual is in training or has been trained—at least for the domain of traits and its elements sampled in this study. In particular, the notion that occupational images converge over time or with experience to some sort of norm—which in this study was assumed to be the image held by the professional group—appears untenable. The number and kinds of traits and their patterning which characterize the average freshman's occupational image, regardless of title, are not much different, at least statistically, from those which characterize the professional's image. On this basis it seems reasonable to conclude that the average college freshman beginning his occupationally relevant

education, his college major, possesses a fairly accurate image—assuming the professionals know what they are talking about—of the typical incumbent of the intended occupation.

Despite some overlap, the common images of the two occupational titles studied were noticeably different in certain respects. First, the industrial management image appeared more extensive and complex than that of electrical engineering. The industrial management image, aside from including the ability, training, and performance factors found in the electrical engineering image, stressed the importance of the personality and interpersonal characteristics of the individual. To perform successfully in this area the individual must show personality traits and interpersonal skills which would permit him to deal effectively with and perhaps influence other people. He must be sensitive to the motives or needs of others, accurately interpret their behavior, and exhibit self-behaviors that do not conflict with his evaluation of these motives and behaviors of others. Interestingly enough, guile or duplicity as an interpersonal strategy—as found in the so-called Machiavellian personality—is not rejected, nor is it accepted, as part of the industrial management image. As would be expected with such an interpersonal orientation, verbal skills were highly valued.

Second, the pattern of importance of traits common to both groups was different. The electrical engineering groups assigned higher importance to intellectual and cognitive skills and their utilization, for example, educational training, ingenuity, inquisitiveness, and flexibility of thought, than did the industrial management Ss. Again, this reflects a tendency for the electrical engineers to be oriented more toward individual or problem-solving activities than the industrial management groups. For the electrical engineers it is sufficient to be bright, well-schooled, ingenious, and persistent in one's approach to problems. Any success accruing to the individual will come from these characteristics rather than from the interpersonal sphere.

As to self-image and its relation to occupational image, on the basis of the large discriminant dispersions, considerable individual differences in self-description were noted. It is

likely that the patterns or profiles of these self-characterizations for each person also show considerable individual differences. If so, it may be possible to break down the total sample of Ss engaging in self-rating into subsets of Ss internally homogeneous with respect to profile, that is, sets which represent distinct types of people. Despite these large observed and hypothesized individual differences, the correlation values obtained in this study indicate that the freshman self-descriptions are similar to the common images of the occupational titles. It may be that this correspondence between self and occupational images is due to common variance attributable to social desirability, a point considered next. In any case, there appear to be marked differences between the ways college freshmen enrolling in electrical engineering and industrial management view and describe themselves, with these two self-images being highly correlated with the corresponding occupational images. Because of these differential correlations it is unlikely that all the correspondence between self- and occupational-images can be accounted for in terms of SD.

The significant correlations between both SD ratings and self-descriptions, and SD ratings and occupational descriptions, indicate that the kind and pattern of traits considered descriptive of self and other in this case correspond to what Ss also consider to be desirable behavior in general. The work of Jackson and Singer (1967) suggests that the notion of SD is indeed a complex one being influenced—among other things—by content of the items, sex of Ss, and individual differences. More importantly, they have demonstrated SD to be a multidimensional construct. These results again indicate the need for considering individual points of view or types in the study of self- or other-descriptions, and, perhaps, the inappropriateness of averaging SD ratings over a sample of Ss (Tucker & Messick, 1963). In this study, however, interest and, thus, conclusions were limited to the gross association between the item parameter of SD and person ratings. It is apparent from the results that much of the self- and other-characterizations studied can be described in terms of gross judgments of socially desirable behaviors. That this finding is not adequate for

completely explaining self and occupational images has been noted. A reasonable conclusion in this respect is that S's perceptions of himself and the typical incumbent of a selected occupation frequently reflect socially desirable behaviors, but that it is unlikely that these perceptions are solely determined by socially desirable response tendencies.

An important question originally set out to be answered concerned the extent to which occupational and self-images are related to vocational behaviors, such as vocational choice, satisfaction, or performance. On the basis of the present results the answer seems to be "not much." For one thing, occupational images apparently form quite early, probably in high school or earlier, and are accurate in the sense of matching the professional's images. Since the students' average self-image tends also to coincide with the occupational norm of the major in which they are enrolled, if the self-concept model is accurate, little dissatisfaction on the part of the students with a current vocational choice, little attraction to another vocational area, or little vocational uncertainty can be expected. Unfortunately, our knowledge relating to college change of major forces us to reject this expectation (Marks, Webb, & Strickland, 1967).

One possible factor in such observed disenchantment with or change in vocational preference or choice among college students is the notion that the students themselves are changing or, in terms of the present development, that their self-image is changing (Plant, 1962). Since the occupational images of a given student probably remain fixed, this hypothesized change in self-image implies a reordering of occupational preferences by the individual thus leading to a change in his choice of vocation and major. Under this interpretation the self-concept model is consistent and can account for vocational behavior and choice.

An alternative position is that trait descriptions, either self- or other-, are not specific enough to permit adequate prediction of vocational preferences or behaviors. For example, simply describing a person as bright, aggressive, emotionally stable, etc., may not be sufficiently precise to yield sensitive or

accurate predictions of that individual's vocational behavior. Although self- and other-images, as studied here, might be valuable adjuncts, the study of vocational preference, choice, and behavior might profitably be extended to include rather specific vocational or job activities. Here a college student might also be asked to express his knowledge of specific job characteristics, activities, or tasks, for example, what does the typical occupational incumbent do both on and off the job. These approaches—the self-concept model and knowledge of specific job activities and behaviors—when taken together, should provide a more rigorous examination and prediction of occupational criteria.

REFERENCES

BOCK, D. R., & JONES, L. V. The measurement and prediction of judgmental response—statistical methods. Psychometric Laboratory, University of North Carolina, 1963.

BORDIN, E. S. A theory of interests as dynamic phenomena. *Educational and Psychological Measurement*, 1943, 3, 49–66.

EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

JACKSON, D. N., & SINGER, J. E. Judgments, items,

and personality. *Journal of Experimental Research in Personality*. 1967, 2, 70–79.

MARKS, E., WEBB, S. C., & STRICKLAND, J. A. Intra-institute transfer report. II: Actual change of major and its reported causes. Research Memorandum 67–8, Evaluation Studies, Georgia Institute of Technology, December 1967.

MARKS, E. Some profile methods useful with singular covariance matrices. *Psychological Bulletin*, 1968, 70, 179–184.

NORMAN, W. T. On estimating psychological relationships: Social desirability and self-report. *Psychological Bulletin*, 1967, 67, 273–293.

O'DOWD, D. D., & BEARDSLEE, D. C. College student images of a selected group of professions and occupations. Cooperative Research Project No. 562 (8143), Wesleyan University, 1960.

PLANT, W. T. *Personality changes associated with a college education*. Cooperative Research Project No. 348, Final Report, 1962, San Jose State College, California.

SUPER, D. E., STARISHEVSKY, R., MATLIN, N., & JORDAAN, J. P. *Career development: Self-concept theory*. New York: College Entrance Examination Board, 1963.

TIEDEMAN, D. V., & O'HARA, R. P. *Career development: Choice and adjustment*. New York: College Entrance Examination Board, 1963.

TORGERSON, C. S. *Theory and methods of scaling*. New York: Wiley, 1958.

TUCKER, L. R., & MESSICK, S. An individual differences model for multi-dimensional scaling. *Psychometrika*, 1963, 28, 333–367.

(Received June 6, 1968)

COLOR CODING EFFECTS IN COMPATIBLE AND NONCOMPATIBLE DISPLAY-CONTROL ARRANGEMENTS

GARY K. POOCK¹

Naval Postgraduate School, Monterey, California

An experiment was designed to assess the effect of color coding in compatible and noncompatible display-control arrangements. Forty Ss viewed four arrangements of display-control panels with color coding or no color coding, and with displays and controls arranged in either a compatible or noncompatible arrangement. The Ss' task was to shut off the display as fast as possible for 80 trials. Color coding was more effective when displays and controls were arranged in a noncompatible fashion, and had no effect when display and control were arranged in a compatible manner. The results support the importance of compatibility in display-control location.

Previous researchers have investigated the effects of color coding in various types of visual displays. Green and Anderson (1956) investigated color coding in a visual search task while Jones (1962) provided a survey of previous research by numerous authors and discussed the various parameters important in color coding. Smith (1962, 1963), Smith and Thomas (1964), and Smith, Farquhar, and Thomas (1965) have reported a series of investigations involving color coding in information displays. Their research, along with that of other researchers, indicates potential benefits of color coding depending on its use and the type of display. In another research study by Chapanis and Lockheed (1965), the effectiveness of sensor lines showing linkages between displays and controls was tested to find out if lines drawn from the control to the display provided any difference in time to respond to a display light on display panels up to 10" × 12". They investigated various panel arrangements using the sensor lines and using compatible versus noncompatible arrangements of display and control. Their results suggested it is more important to make the relative location of displays and controls compatible than it is to use lines showing which control relates to which display. When

the displays were compatible, sensor lines contributed nothing to speed of response.

Recognizing the potential uses of color coding, the purpose of this research was to investigate the use of color coding in compatible and noncompatible arrangements of displays and controls. The specific objective was to determine if color coding would provide different times to control response, especially in a noncompatible display-control arrangement.

METHOD

Forty Army, Navy, and Marine male officers with normal color vision served as experimental Ss. The Ss were divided into four groups of 10 each and were tested for 80 trials in one of four display-control arrangements. A partially nested analysis of variance was used, with Ss nested in one of the four display-control arrangements but common to the 80 trials. The compatible and noncompatible arrangements with color coding are shown in Figure 1. The panels were 12" × 24". The displays were standard red, green, blue, and white Christmas tree lights and the controls were standard light switches with mounting plates located directly beneath the lights in a corresponding fashion. The null position of all switches was toward S.

The compatible arrangement with color coding is shown on the left side of Figure 1. The compatible arrangement with no color coding was exactly the same except all lights were red and all switches and mounting plates were white. However, the upper left-hand light was still controlled by the upper left-hand switch and likewise for the other three display controls. In the noncompatible color coded arrangement shown on the right side of Figure 1, a colored switch still controlled its respective colored light, but the location of displays and controls was not compatible. The noncompatible arrangement with

¹ The author expresses his appreciation to Dennis R. Jordan, LCDR, USN, and Gerald A. Vick, Major, USA, for their help in procurement of the apparatus and data. Requests for reprints should be sent to the author, Department of Operations Analysis, Naval Postgraduate School, Monterey, California 93940.

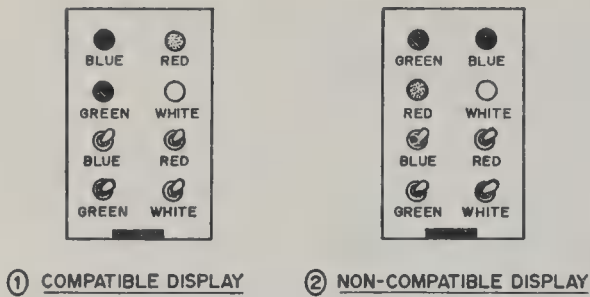


FIG. 1. Color coded compatible and noncompatible displays used. (Lettering is included for easier interpretation and did not actually appear on displays.)

no color code again used red lights and white switches, but *S* had to remember which switch controlled which light. The display-control arrangement was the same as the noncompatible with color.

At the start of a test session, *E* demonstrated the use of the display-control panel an *S* was to use. Standardized instructions informed each *S* that he was to start each trial with his hand on the black marker at the bottom center of the control panel. The *S*'s task would be to shut off a particular light as fast as he could by pushing the corresponding switch away from him. If he pushed an incorrect switch, the light would stay on until the correct switch was activated. After 12 practice trials and any questions, the testing period began. Each light occurred randomly an equal number of times in each half of the trials. Time to correct response was measured in tenths of a second. This parameter had been suggested by Chapanis and Lockhead (1965) as being more meaningful operationally, even though it may be more complex psychologically than time to first response.

RESULTS

Figure 2 shows the average time to correct response for each of the panel combinations

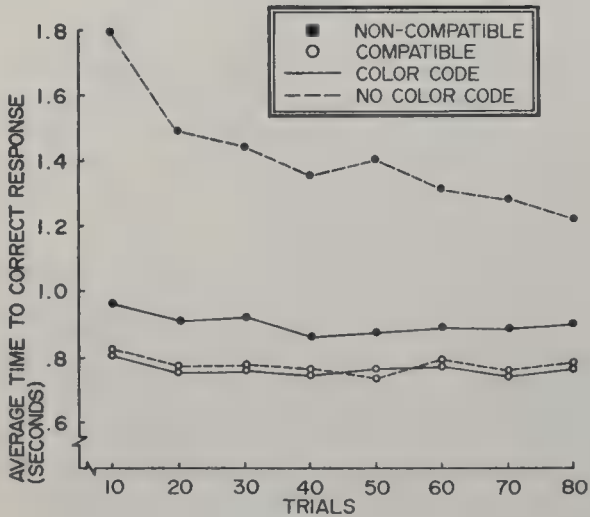


FIG. 2. Average times to correct response. (Each point is average of 100 data points.)

used. The graph is very similar to the corresponding one of Chapanis and Lockhead except the current data provided uniformly higher response times because of the different controls used.

It was desired to examine the effects of color coding and display-control arrangement only after initial learning was apparently over. An overall analysis of variance indicated a significant difference ($p < .005$) between the first and last half of the 80 trials. This result and observation of the data indicated a significant practice effect had occurred. However, further analysis on the second half of the trials indicated no significant improvement in response times. Thus Table 1 shows the analysis of variance on the last 40 trials, and should more reasonably indicate what significant effects remain in the data after the first 40 trials of performance. Therefore, the remainder of the paper will be concerned with Table 1.

The most outstanding result of this investigation was the effect of the combination of color coding and panel arrangement where the very marked beneficial effect of color coding on a spatially noncompatible panel was demonstrated. The reader's attention is directed to the top two curves of Figure 2, where the data of the last 40 trials indicate an improvement in response times of 40–50% when color coding is used on noncompatible panel arrangements. This result pro-

TABLE 1
ANALYSIS OF VARIANCE ON SECOND
HALVES OF THE DATA
(TRIALS 41 THROUGH 80)

| Sources of variation | <i>df</i> | <i>MS</i> | <i>F</i> |
|--|-----------|-----------|----------|
| Between <i>Ss</i> (<i>S</i>) | 39 | 309.15 | — |
| Between panels (<i>P</i>) | 3 | 2729.59 | 25.40* |
| Compatible vs. noncompatible (<i>C</i>) | 1 | 4613.80 | 42.94* |
| Color code vs. no color code (<i>Col</i>) | 1 | 1740.97 | 16.20* |
| Interaction: <i>C</i> × <i>Col</i> | 1 | 1833.99 | 17.07* |
| Between <i>Ss</i> within panels (<i>S</i> w/ <i>P</i>) | 36 | 107.45 | — |
| Between trials within <i>Ss</i> | 1560 | 8.97 | — |
| Total | 1599 | | |

* $p < .001$.

vides strong support for the use of color coding when displays and controls cannot be arranged in a spatially compatible relationship. Color coding had no effect on response times when the displays were compatible, as seen in the two bottom curves in Figure 2. Further, color coding did not improve response levels on noncompatible displays to those of compatible displays.

There was a significant difference between compatible and noncompatible displays with the compatible displays providing the faster response times. The significance of the effect of compatibility is supported by further analysis which showed approximately 55% of the variance between panels was due to the compatibility of the panel arrangement. This strong effect of compatibility is in agreement with Chapanis and Lockhead where they showed 77% of the variance between panels was due to the compatible-noncompatible effect.

The main effect of color coding was also significant and did account for more of the total variance between panels than the sensor lines of Chapanis and Lockhead. Approximately 20% of the variance between panels was due to color coding in this experiment versus approximately 6% due to sensor lines in their experiment.

Finally, further analysis using Duncan's multiple range test (Winer, 1962) for pairwise differences among the four panels showed no significant difference between the compatible panels at the .05 level. All other pairwise differences between the panels were significant ($p < .05$).

DISCUSSION

The compatibility of the display and control with respect to relative spatial location is the most important point revealed in this investigation and confirms that same conclusion of Chapanis and Lockhead (1965). Color coding appears to have a proportionally more significant effect than the use of sensor lines.

The conclusions from this investigation suggest, in view of the panels used here, that one should first concentrate on compatibility with respect to display-control location. However, if this is not possible, the use of color coding will improve significantly the time to correct response in a noncompatible display.

REFERENCES

- CHAPANIS, A., & LOCKHEAD, G. R. A test of the effectiveness of sensor lines showing linkages between displays and controls. *Human Factors*, 1965, 7, 219-230.
- GREEN, B. F., & ANDERSON, L. K. Color coding in a visual search task. *Journal of Experimental Psychology*, 1956, 51, 19-24.
- JONES, M. R. Color coding. *Human Factors*, 1962, 4, 355-365.
- SMITH, S. L. Color coding and visual search. *Journal of Experimental Psychology*, 1962, 64, 434-440.
- SMITH, S. L. Color coding and visual separability in information displays. *Journal of Applied Psychology*, 1963, 47, 358-364.
- SMITH, S. L., & THOMAS, D. W. Color versus shape coding in information displays. *Journal of Applied Psychology*, 1964, 48, 137-146.
- SMITH, S. L., FARQUHAR, B. B., & THOMAS, D. W. Color coding in formatted displays. *Journal of Applied Psychology*, 1965, 49, 393-398.
- WINER, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

(Received June 26, 1968)

SVIB INTERESTS AND PRODUCT PREFERENCES

ROBERT A. PETERSON¹ AND ALLAN L. PENNINGTON

Graduate School of Business Administration, University of Minnesota

A study was conducted to relate interest patterns and consumer product preferences. The SVIB and a product preference questionnaire were administered to a sample of 239 male business students. Significant interest differences were found between Ss preferring a convertible automobile to a sedan automobile, a trip to Yellowstone to a trip to Las Vegas, and a savings account to common stock. These differences were aligned on a people-nonpeople interest axis, with the latter product in each pair being associated with relatively more people-oriented interest patterns, and suggest that meaningful interest-product preference relationships may exist.

Numerous attempts have been made to relate various psychological concepts to consumer purchasing patterns. Examples are the studies of Wells (1961) in predicting consumer behavior by the use of attitudes; Krugman and Hartley (1960) in the learning of tastes; Tucker and Painter (1961) and Westfall (1962) in correlating personality factors with product preferences and usage rates; and Losciuto and Perloff (1967) in relating cognitive dissonance to consumer preferences.

However, no attempts have been made to relate interest patterns to consumer product preferences. This is surprising since interests are viewed as "complex psychological structures controlling choices of the use one will make of his time" (Tyler, 1965, p. 206); hence, interest patterns should be related to decisions to purchase several types of consumer products.

To investigate these relationships, two pilot studies were undertaken using the Strong Vocational Interest Blank (SVIB) and a product preference questionnaire. These studies revealed that scores on SVIB physical science scales (Group II), technical and skilled trade scales (Group IV), social service scales (Group V), and business and sales scales (Groups VIII and IX) differentiated between respondents who preferred certain products presented in a forced-choice situation.

The primary purpose of the present study was to replicate these pilot studies and cross-

validate preliminary findings with a similar and larger sample. A second purpose was to suggest the potential of previously unexplored psychological concepts as determinants of consumer behavior.

METHOD

Subjects

The sample consisted of 239 male students enrolled in the introductory marketing course at the University of Minnesota. The mean age was 21.8 yr. with an SD of 1.5 yr.

Instruments

The instruments used were the SVIB and a product preference questionnaire. The SVIB was used because it has consistently proved to be valid and reliable in vocational counseling and personnel selection for over four decades (Campbell, 1966). Furthermore, it has been successfully applied in several other contexts (Campbell & Johansson, 1966; Knapp, 1964; Thorndike, Weiss, & Dawis, 1968; Whitehorn & Betz, 1960).

The product questionnaire consisted of structured questions concerning product preferences. Products were presented in pairs on a 5-point scale ranging from strongly preferring Product A to strongly pre-

TABLE 1
PRODUCT ALTERNATIVES AND NUMBER OF SUBJECTS
PREFERRING EACH PRODUCT IN A PAIR

| Product pair | Number of Ss preferring |
|--|-------------------------|
| Savings account Common stock | 73 144 |
| Convertible automobile Sedan automobile | 104 106 |
| Yellowstone trip Las Vegas trip | 65 135 |

¹ Requests for reprints should be sent to Robert A. Peterson, School of Business Administration, University of Minnesota, Minneapolis, Minnesota 55455.

TABLE 2

SVIB MEANS AND STANDARD DEVIATIONS FOR SAVINGS ACCOUNT PREFERRERS ($n = 73$)
AND COMMON STOCK PREFERRERS ($n = 144$)

| Scale | Savings preferers | | Stock preferers | | D_m |
|--------------------------------------|-------------------|------|-----------------|------|---------|
| | M | SD | M | SD | |
| Group II Architect | 22.0 | 9.2 | 18.5 | 9.5 | 3.5* |
| Mathematician | 13.6 | 7.9 | 10.2 | 9.0 | 3.4** |
| Physicist | 12.8 | 9.5 | 9.6 | 10.8 | 3.2* |
| Chemist | 19.3 | 11.6 | 17.0 | 13.3 | 2.3 |
| Engineer | 22.7 | 10.7 | 20.3 | 11.1 | 2.4 |
| Group IV Carpenter | 21.3 | 11.5 | 16.8 | 12.2 | 4.5** |
| Forest service man | 20.0 | 12.8 | 19.4 | 11.9 | .6 |
| Farmer | 29.9 | 10.2 | 28.1 | 9.4 | 1.8 |
| Math-science teacher | 30.7 | 10.4 | 29.6 | 11.1 | 1.1 |
| Printer | 34.1 | 9.5 | 30.0 | 9.8 | 4.1** |
| Policeman | 22.3 | 9.3 | 20.7 | 9.4 | 1.6 |
| Group V Personnel manager | 30.3 | 9.3 | 35.1 | 10.6 | -4.8*** |
| Public administrator | 34.8 | 9.0 | 38.4 | 10.7 | -3.6* |
| Rehabilitation counselor | 32.8 | 10.1 | 34.0 | 10.4 | -1.2 |
| YMCA secretary | 33.8 | 11.1 | 37.6 | 12.4 | -3.8* |
| Community recreation director | 34.0 | 10.7 | 37.5 | 12.4 | -3.5* |
| Social worker | 28.0 | 12.8 | 30.9 | 12.8 | -2.9 |
| School superintendent | 19.2 | 9.9 | 20.9 | 11.1 | -1.7 |
| Minister | 12.4 | 14.0 | 12.1 | 13.7 | .3 |
| Group IX Sales manager | 35.3 | 10.5 | 39.0 | 11.4 | -3.7* |
| Real estate salesman | 39.2 | 8.7 | 40.7 | 8.5 | -1.3 |
| Life insurance salesman | 31.7 | 9.9 | 34.6 | 10.7 | -2.9 |
| Chamber of commerce executive | 39.2 | 8.6 | 44.2 | 10.4 | -5.0*** |
| Credit manager | 40.9 | 9.5 | 44.9 | 9.5 | -4.0** |
| Occupation introversion-extraversion | 44.7 | 11.2 | 39.7 | 11.0 | 5.0** |
| Occupational level | 55.4 | 6.4 | 58.2 | 8.1 | -2.8* |
| President, manufacturing firm | 26.4 | 8.5 | 29.5 | 9.5 | -3.1* |
| Adventure ^a | 60.6 | 9.9 | 63.9 | 9.1 | -3.3* |
| Public speaking ^a | 52.1 | 10.0 | 55.7 | 10.0 | -3.6* |

^a Basic scales.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

ferring Product B. A sixth category was provided for "prefer neither" responses. Criteria used in selecting product pairs were (a) preferences for particular products were related to specific interest patterns in the pilot studies; (b) product pairs represented a variety of consumer alternatives—travel, investment, a major durable good purchase; and (c) the products were feasible alternatives and similar in price. Undesirable product pairs were eliminated by pretests; only those meeting the above criteria were included in this study.

Procedure

The SVIB and product questionnaire were administered as unrelated instruments in a single testing session; thus, students were led to believe they

were participating in two separate studies. A subsequent check indicated that the participants did not associate the two instruments.

Three pairs of product alternatives, listed in Table 1, were investigated by comparing mean SVIB scores of *Ss* strongly or moderately preferring one product with the mean scores of those *Ss* preferring the other. "Indifferent" and "prefer neither" categories were excluded from the analysis. Table 1 also presents the number of *Ss* strongly or moderately preferring each alternative.

RESULTS

Tables 2, 3, and 4, respectively, report selected SVIB scale means and standard devia-

tions for the analyzed pairs of products. In general, the results of the present study confirmed those of the pilot studies. Specifically:

- (1) The Ss preferring savings accounts obtained higher scores on SVIB physical science and technical-skilled trade scales (Groups II and IV) and lower on social service, business, and sales scales (Groups V, VIII, and IX) than Ss preferring common stock.
- (2) The Ss preferring a trip to Yellowstone scored higher on technical-skilled trade scales

- (Group IV) and lower on sales and verbal-linguistic scales (Groups IX and X) than Ss preferring a trip to Las Vegas.
 - (3) The Ss preferring a convertible automobile obtained higher mean scores on social service and sales scales (Groups V and IX) and lower scores on physical science and technical-skilled trade scales (Groups II and IV) than Ss preferring a sedan automobile.
- Only those SVIB scale groups differentiating between subsamples in the pilot studies

TABLE 3
SVIB MEANS AND STANDARD DEVIATIONS FOR CONVERTIBLE PREFERRERS (*n* = 104)
AND SEDAN PREFERRERS (*n* = 106)

| Scale | | Convertible preferers | | Sedan preferers | | D _m |
|----------|--------------------------------------|-----------------------|-----------|-----------------|-----------|----------------|
| | | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | |
| Group II | Architect | 17.2 | 9.7 | 21.4 | 8.2 | −4.2*** |
| | Mathematician | 8.0 | 8.2 | 14.0 | 8.3 | −6.0*** |
| | Physicist | 7.4 | 10.2 | 13.1 | 10.1 | −5.7*** |
| | Chemist | 14.9 | 12.9 | 19.6 | 12.0 | −4.7** |
| | Engineer | 18.4 | 11.5 | 23.7 | 10.3 | −3.3*** |
| Group IV | Carpenter | 15.6 | 12.0 | 21.1 | 11.9 | −5.5*** |
| | Forest service man | 19.0 | 12.0 | 20.6 | 12.0 | −1.6 |
| | Farmer | 26.6 | 10.0 | 31.2 | 9.3 | −4.6*** |
| | Math-science teacher | 29.1 | 10.2 | 30.5 | 11.2 | −1.4 |
| | Printer | 29.9 | 10.1 | 33.2 | 9.6 | −3.3* |
| | Policeman | 21.9 | 9.3 | 21.0 | 9.3 | .9 |
| Group V | Personnel director | 34.5 | 11.1 | 33.2 | 10.4 | 1.3 |
| | Public administrator | 38.1 | 10.7 | 36.3 | 10.2 | 1.8 |
| | Rehabilitation counselor | 35.5 | 10.7 | 31.6 | 9.7 | 3.9** |
| | YMCA secretary | 40.3 | 12.4 | 32.3 | 10.7 | 8.0*** |
| | Social worker | 32.3 | 13.2 | 27.3 | 12.4 | 5.0** |
| | Social science teacher | 36.1 | 11.0 | 32.9 | 9.5 | 3.2* |
| | School superintendent | 21.2 | 10.9 | 19.6 | 10.5 | 1.6 |
| | Minister | 12.9 | 13.9 | 10.7 | 14.0 | 2.2 |
| Group IX | Sales manager | 40.3 | 10.0 | 36.6 | 11.2 | 3.7* |
| | Real estate salesman | 41.5 | 7.9 | 39.5 | 8.4 | 2.0 |
| | Life insurance salesman | 35.8 | 9.9 | 31.9 | 10.0 | 3.9** |
| | Biologist | 10.0 | 9.3 | 14.5 | 11.0 | −4.5** |
| | Chamber of commerce executive | 45.2 | 10.2 | 39.9 | 9.3 | 5.3*** |
| | Community recreation director | 39.7 | 12.0 | 32.8 | 10.6 | 6.9*** |
| | Occupation introversion-extraversion | 38.4 | 11.3 | 44.8 | 10.7 | −6.4*** |
| | Sales ^a | 59.3 | 8.6 | 55.7 | 9.0 | 3.6** |
| | Adventure ^a | 64.9 | 8.1 | 60.6 | 10.1 | 4.3*** |
| | Recreational leadership ^a | 56.9 | 8.1 | 52.3 | 9.2 | 4.6*** |
| | Public speaking ^a | 55.7 | 9.9 | 53.0 | 9.9 | 2.7* |
| | Merchandising ^a | 63.1 | 6.7 | 60.5 | 7.1 | 2.6** |

^a Basic scales.
* *p* < .05.
** *p* < .01.
*** *p* < .001.

TABLE 4

SVIB MEANS AND STANDARD DEVIATIONS FOR TRIP TO YELLOWSTONE PREFERRERS ($n = 65$)
AND TRIP TO LAS VEGAS PREFERRERS ($n = 135$)

| Scale | Yellowstone preferrers | | Las Vegas preferrers | | D_m |
|--------------------------------------|------------------------|------|----------------------|------|---------|
| | M | SD | M | SD | |
| Group IV Carpenter | 21.8 | 11.1 | 16.7 | 12.4 | 5.1** |
| Forest service man | 24.4 | 12.2 | 17.3 | 11.5 | 7.1*** |
| Farmer | 31.1 | 9.8 | 27.6 | 9.8 | 3.5* |
| Math-science teacher | 32.9 | 10.8 | 28.4 | 11.0 | 4.5** |
| Printer | 32.1 | 10.6 | 31.5 | 9.8 | .6 |
| Policeman | 23.5 | 9.3 | 20.3 | 9.1 | 3.2* |
| Group IX Sales manager | 34.3 | 11.2 | 40.2 | 10.4 | -5.9*** |
| Real estate salesman | 38.0 | 8.7 | 41.7 | 8.2 | -3.7** |
| Life insurance salesman | 30.5 | 10.2 | 35.8 | 9.9 | -5.3*** |
| Group X Advertising man | 28.1 | 10.0 | 33.6 | 10.3 | -5.5*** |
| Lawyer | 26.4 | 8.1 | 29.2 | 8.4 | -2.8* |
| Author-journalist | 25.1 | 6.5 | 28.2 | 7.4 | -3.1** |
| Biologist | 15.8 | 10.4 | 10.3 | 10.0 | 5.5*** |
| Chamber of commerce executive | 39.7 | 9.2 | 44.2 | 10.1 | -4.5** |
| Liberalism-conservatism | 41.3 | 9.3 | 45.4 | 9.3 | -4.1** |
| President, manufacturing firm | 26.8 | 9.1 | 29.4 | 8.6 | -2.6* |
| State department interpreter | 25.1 | 11.6 | 31.2 | 11.2 | -6.1*** |
| Mechanical ^a | 49.6 | 9.3 | 44.9 | 9.1 | 4.7*** |
| Nature ^a | 46.5 | 10.5 | 39.2 | 8.7 | 7.3*** |
| Adventure ^a | 60.4 | 10.5 | 64.1 | 8.9 | -3.7* |
| Recreational leadership ^a | 52.9 | 9.1 | 55.7 | 8.3 | -2.8* |

^a Basic scales.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

are reported in the tables. However, all of the individual scales within the reported groups are presented to illustrate the consistency of the results.

In addition, selected Occupational, Nonoccupational, and Basic Interest scales (Campbell et al., 1968) are presented at the bottom of each table. These supplemented the preceding grouped scales and assisted in the interpretation of the results.

DISCUSSION AND CONCLUSIONS

A content analysis of the interests of respondents preferring different alternatives suggest these interests can be associated with a people-nonpeople interest continuum. The Ss who preferred savings accounts, trip to Yellowstone, and sedan automobile reported more nonpeople interests, while Ss who preferred

common stock, trip to Las Vegas, and convertible reported more people-oriented interests. The supplemental scales reported at the bottom of Tables 1-3 support this contention. Particularly noteworthy are the respective scores on Occupational introversion-extraversion, a direct measure of people interests (lower scores) versus nonpeople interests (higher scores), and the Public speaking basic scale. In two of the product pairs, respondents on the people-end of the interest continuum scored significantly lower on the former scale and higher on the latter scale, and in the third pair the mean differences were in the expected directions.

Moreover, the supplemental scales lend face validity to the study findings. For instance, Ss who preferred savings accounts scored significantly lower on the Occupational level scale than Ss who preferred common stock.

These scores illustrate the lower socioeconomic interests of Ss who prefer savings accounts and are consistent with the research findings of Katona (1964). Additionally, the largest mean score difference between Ss who preferred a trip to Yellowstone and those who preferred a trip to Las Vegas was on, appropriately, the Nature scale. Finally, Ss preferring common stock, the Las Vegas trip, and convertible automobile obtained significantly higher scores on the Adventure scale. The differences on this scale were expected from the nature of the product alternatives and add credence to the concept of an interest-consumer preference relationship.

Although the results must be interpreted with respect to the sample employed, the interdependence of the interest score differences, and the specific products used, they nevertheless suggest that meaningful relationships may exist between interest patterns and product preferences. In fact, the interest homogeneity of the sample and the methodology employed probably increase the practical significance of the results.

Given that these results can be replicated across random consumer samples with actual purchases as the criteria, interest differences may prove to be much greater. If this is so, interest patterns may ultimately become important variables in the explanation and prediction of purchasing behavior. Further research should focus upon two areas: (a) Interests should be investigated with respect to the amount of variance they account for in actual purchase decisions compared with other purchase determinants, and (b) attempts

should be made to analyze similarities of interest patterns across product categories.

REFERENCES

- CAMPBELL, D. P. *Manual for the Strong Vocational Interest Blank*. Palo Alto: Stanford University Press, 1966.
- CAMPBELL, D. P., & JOHANSSON, C. B. Academic interests, scholastic achievements, and eventual occupations. *Journal of Counseling Psychology*, 1966, 13, 416-424.
- CAMPBELL, D. P., BORGES, F. H., EASTES, S. H., JOHANSSON, C. B., & PETERSON, R. A. A set of basic interest scales for the Strong Vocational Interest Blank for Men. *Psychological Monographs*, December, 1968.
- KATONA, G. *The mass consumption society*. New York: McGraw-Hill, 1964.
- KNAPP, R. H., & HOFFMAN, E. G. Vocational interest scores and patterns of aesthetic preference. *Journal of Social Psychology*, 1964, 64, 299-306.
- KRUGMAN, H. E., & HARTLEY, E. L. The learning of tastes. *The Public Opinion Quarterly*, 1960, 24, 621-631.
- LOSCIUTO, L. A., & PERLOFF, R. Influences of product preference on dissonance reduction. *Journal of Marketing Research*, 1967, 4, 286-290.
- THORNDIKE, R. M., WEISS, D. J., & DAWIS, R. V. Canonical correlation of vocational interests and vocational needs. *Journal of Counseling Psychology*, 1968, 15, 101-106.
- TUCKER, W. T., & PAINTER, J. J. Personality and product use. *Journal of Applied Psychology*, 1961, 45, 325-329.
- TYLER, L. E. *The psychology of human differences*. New York: Appleton-Century-Crofts, 1965.
- WELLS, W. D. Measuring readiness to buy. *Harvard Business Review*, 1961, 39, 81-87.
- WESTFALL, R. Psychological factors in predicting product choice. *Journal of Marketing*, 1962, 26, 34-40.
- WHITEHORN, S. C., & BETZ, B. J. Further studies of the doctor as a crucial variable in the outcome of treatment with schizophrenic patients. *American Journal of Psychiatry*, 1960, 117, 215-223.

(Received July 8, 1968)

EFFECTIVENESS OF SVIB ACADEMIC INTEREST SCALES IN PREDICTING COLLEGE ACHIEVEMENT¹

RICHARD W. JOHNSON²

Counseling Center, University of Wisconsin

The effectiveness of various SVIB academic interest scales in predicting first semester grades for freshman males at the University of Massachusetts was determined. Both the Rust and Ryan and the Campbell and Johansson scales contributed significantly, albeit modestly, to a multiple correlation coefficient consisting of high school rank and SAT scores in predicting academic performance. A single-item, self-evaluation rating scale failed to predict grade point average significantly. Although the degree of relationship between the interest scales and grades tended to be somewhat greater for "marginal" students, the r 's were not significantly different from those obtained with more able students. The use of modified, "placement" instructions did not greatly affect the mean scores or the magnitude of the correlations.

In recent years, several new scales have been developed from Strong Vocational Interest Blank (SVIB) items to aid both in predicting college achievement and in understanding the motivational and temperamental factors associated with academic success (Campbell & Johansson, 1966; Martin, 1964; Rust & Ryan, 1954). Each of these scales has survived cross-validation study in at least one setting.

Rust and Ryan (1954) developed separate scales to predict overachievement, normal achievement, and underachievement at Yale University. These scales significantly differentiated between various groups of over-, normal, and underachievers at both Yale University and, more recently, Harvard University (McArthur, 1965).

Martin (1964) constructed a series of academic interest scales (both long and short forms) from SVIB items for males and females enrolled in liberal arts and males enrolled in engineering and mines at the University of Pittsburgh. The scales significantly contributed to a multiple correlation based in part upon Scholastic Aptitude Test (SAT) scores and high school rank in predicting first year grades for succeeding classes at the University of Pittsburgh.

¹ The research reported herein was performed pursuant to a contract with the Office of Education of the United States Department of Health, Education and Welfare while the author was on the staff at the University of Massachusetts.

² Requests for reprints should be sent to the author, Counseling Center, University of Wisconsin, 415 Gilman Street W., Madison, Wisconsin 53706.

Finally, Campbell and Johansson (1966) developed the Academic Achievement (AACH) scale to differentiate between high and low achievers in the College of Liberal Arts at the University of Minnesota. Although the AACH scale correlated significantly with first year grade point average in the cross-validation study at Minnesota ($r = .36$), it did not significantly contribute to a multiple correlation coefficient consisting of a scholastic aptitude test and high school rank in predicting grades. The scale provides some insights into the personal, motivational characteristics associated with high and low grades. This new nonoccupational interest scale has been added to the profile of the 1966 revision of the SVIB (Strong & Campbell, 1966).

Will these various interest scales be effective in predicting achievement in a new academic setting? The present study was primarily addressed to this question. Four specific questions concerning the practical application of the academic interest scales were asked:

1. Are the SVIB scales more effective than a single-item, self-rating scale? Holland and Lutz (1968), in particular, have argued that simple, direct questions might produce results as effective as, or possibly more effective than, long lists of inventory items.

2. Are the SVIB academic interest scales more highly correlated with college grades for "marginal" students than for superior or average students? Clark (Clark, 1961; Clark & Campbell, 1965) has presented data suggest-

ing that when learning ability was "just adequate," the correlation between interests and achievement was more pronounced.

3. Is the degree of relationship affected if the students take the inventory with instructions that the results may be used for placement purposes? While it has been long known that the SVIB profile may be rather easily faked, the effect of such distortion is not well established. Some studies (e.g., Ruch & Ruch, 1967) suggest that "real life" incentives to fake may actually improve the predictive validities of inventories. If the *S* knows for what purposes the tests will be used, he will be better able to indicate the specific role which he is willing to play in that particular situation (Hathaway, 1960).

4. Do the SVIB scales significantly contribute to a multiple correlation coefficient based in part upon SAT scores and high school rank in predicting college achievement? To be most helpful in both counseling and selection, the interest scales should aid in accounting for that part of the variance in college grades not already accounted for by readily available intellectual predictors.

METHOD

Subjects

The final sample consisted of 290 freshman males enrolled in the College of Arts and Sciences (A&S) and 98 freshman males enrolled in the School of Business Administration (SBA) at the University of Massachusetts who participated in the summer orientation program in 1967. Eleven of the A&S students and two of the SBA students originally tested were not included in the final sample because of failure to enroll in college, failure to complete the first semester, or lack of SAT scores.

Measuring Instruments

Each student was asked to complete the SVIB and a single item, self-evaluation rating scale as part of the regular precollege testing program administered during freshman summer orientation sessions. The self-evaluation rating scale consisted of a single, horizontal line with 100 percentage points drawn on it. The student was instructed to mark as closely as possible the percentage point which best represented the percent of other freshman male students whom he felt that he would surpass in terms of his first semester grade point average.

The SVIB was scored to yield the following seven academic interest measures: (1-4) Overachievers, Normalachievers, Underachievers, Overachievers mi-

nus Underachievers (O minus U) (Rust & Ryan, 1954); (5-6) Academic Interest Scale (AIS): Liberal Arts Males (LAM), 1959 version, Long and Short Forms (Martin, 1964), and (7) Academic Achievement Scale (Campbell & Johansson, 1966).

The 1966 revision of the SVIB (Strong & Campbell, 1966) was used. As 109 of the 400 items on the SVIB were dropped, both the Rust and Ryan and the Martin scales, which were based on the old form of the SVIB, have fewer items on the new form. The effect of this reduction in the number of items on the intercorrelation of the old form with the new form and upon test-retest reliability was determined by means of a sample of 101 young adults who took the old form of the SVIB (which includes all the items scored on the new form) twice over a 30-day interval.³

The instructions for both the SVIB and the self-rating scale were modified for one-fourth of the total sample. The modified instructions informed the students that the results might be used in placing them in advanced courses. The specific instructions are given below. The routine SVIB instructions read as follows: "Among other things, research has shown that this test is helpful in making vocational and educational plans. The test enables the student to compare his interests with those of people employed in various occupations. High scores indicate occupational similarity; low scores indicate dissimilarity. The test results serve as an index of the type of work which you will find interesting. The results will be used in *discussing* occupational and educational plans with you."

The modified SVIB instructions read as follows: "Among other things, research has shown that this test is a fairly good index of academic motivation. Students who receive high 'academic motivation' scores generally do well in their college courses. Students who obtain low 'academic motivation' scores often experience difficulty in their courses. The test results may serve as a measure of your motivation or desire to do well in your course work. As such, the results may be used to guide your *placement* in some of our challenging courses."

Similarly, the instructions for the self-rating scale were varied. The routine instructions began as follows: "Your estimate of your first semester academic performance will be helpful to your counselor in *discussing* your program of courses with you."

The modified instructions began as follows: "Your estimate of your first semester performance will be used as an index of your desire to do well in your course work. As such, it may be used as a guide in *placing* you in some of our more challenging courses."

The tests were administered to the students in groups of 30 to 60 by University personnel. Every fourth test folder contained the modified instructions. The sections with the modified instructions were not read aloud. Of the 290 A&S students in-

³ These data were generously supplied by David P. Campbell, Director, Center for Interest Measurement Research. The composition of the sample is described elsewhere (Strong & Campbell, 1966, p. 27).

TABLE 1
TEST-RETEST RELIABILITIES OF REVISED AND OLD FORMS OF SVIB
ACADEMIC INTEREST MEASURES

| Scale | No. of items | Test | | Retest | | Test-retest reliability | Intercorrelation of revised and old forms |
|-----------------------------------|--------------|----------|-----------|----------|-----------|-------------------------|---|
| | | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | |
| Rust and Ryan Scales ^a | | | | | | | |
| Underachievers | | | | | | | |
| Revised | 37 | -2.6 | 4.2 | -2.9 | 4.2 | .67 | .98 |
| Old | 43 | -2.9 | 4.3 | -3.3 | 4.6 | .68 | |
| Normalachievers | | | | | | | |
| Revised | 24 | .7 | 3.6 | 1.1 | 3.8 | .72 | .96 |
| Old | 29 | .5 | 4.2 | 1.1 | 4.3 | .72 | |
| Overachievers | | | | | | | |
| Revised | 23 | 2.1 | 3.3 | 1.5 | 3.2 | .74 | .93 |
| Old | 34 | 3.1 | 3.9 | 2.2 | 4.0 | .77 | |
| O minus U | | | | | | | |
| Revised | | 4.8 | 5.8 | 4.4 | 5.7 | .72 | .97 |
| Old | | 6.1 | 6.3 | 5.5 | 6.3 | .72 | |
| Martin Scales (LAM) ^b | | | | | | | |
| AIS-Short | | | | | | | |
| Revised | 24 | 13.7 | 2.9 | 13.5 | 2.9 | .77 | .96 |
| Old | 31 | 17.9 | 3.4 | 17.6 | 3.3 | .74 | |
| AIS-Long | | | | | | | |
| Revised | 83 | 46.9 | 6.0 | 45.9 | 6.5 | .76 | .97 |
| Old | 107 | 60.1 | 7.1 | 58.2 | 7.6 | .76 | |
| AACH Scale ^c | 55 | 48 | 12 | 47 | 12 | .88 | |

Note.—*n* = 101.

^a Responses weighted -1, 0, or +1.

^b Responses weighted 0 or 1.

^c Data from SVIB Manual (Strong & Campbell, 1966).

cluded in the final sample, 68 received the modified instructions. Of the 98 SBA students in the final sample, 26 received the modified instructions.

Predicted grade point averages (PGPA), obtained by means of multiple regression equation based upon Converted Class Rank (secondary school rank) and SAT Verbal and Mathematics scores were obtained for all Ss from the Dean of Admissions Office (Glover, 1963). The current version of this formula for freshman males enrolled in either Arts and Sciences or Business Administration is: $PGPA = .01 \text{ SAT-Verbal} + .038 \text{ Converted Class Rank} - .545$. Both SAT-Verbal and Converted Class Rank are expressed as T scores with an *M* of 50 and an *SD* of 10.

Data Analysis

The relationship between the academic interest scales and first semester grade point average (GPA) was determined for the following groups of Ss: A&S students (routine instructions), SBA students (routine instructions), A&S and SBA students (routine instructions), A&S and SBA students (modified instructions), and high-, middle-, and low-predicted GPA groups.

The A&S and SBA students were combined in some instances to increase the size of *n*. This combination appears to be justified in that the two groups of

students were enrolled in essentially the same program of courses for the first semester.

The A&S students were divided as equally as possible into three levels of PGPA. High-predicted GPA, "superior" student ($PGPA = 2.3$ or higher; $n = 81$), middle-predicted GPA "average" student ($PGPA = 2.1$ or 2.2 ; $n = 85$), and low-predicted GPA, "marginal" student ($PGPA = 2.0$ or lower; $n = 56$) groups were formed.

Both zero order and multiple correlation coefficients were computed. The significance of the increase in multiple *R* due to the inclusion of additional variables was tested by means of the analysis of variance procedure described in McNemar (1962, p. 284).

RESULTS

The use of the revised version of the SVIB did not appreciably affect the relative scores of Ss (see Table 1). The intercorrelation of the old and the revised forms was in no case less than .93. The test-retest reliabilities were virtually the same for both the old (long) and the revised (short) forms of the scales.

The means and standard deviations of variables for all groups of Ss are shown in

TABLE 2
MEANS AND STANDARD DEVIATIONS OF VARIABLES FOR ALL GROUPS

| Variable | | Arts and Sciences | School of Business Administration | Groups | | | | |
|-------------------------------------|-----------|-------------------|-----------------------------------|----------------|---------------|---------------|---------------|---------------|
| | | | | A & S and SBA | Place-ment | High PGPA | Middle PGPA | Low PGPA |
| | | <i>n</i> = 222 | <i>n</i> = 72 | <i>n</i> = 294 | <i>n</i> = 94 | <i>n</i> = 81 | <i>n</i> = 85 | <i>n</i> = 56 |
| Self-rating scale | <i>M</i> | 68.1 | 63.9 | 67.1 | 68.3 | 72.2 | 66.2 | 64.9 |
| | <i>SD</i> | 12.2 | 12.2 | 12.3 | 11.9 | 11.4 | 11.9 | 12.4 |
| Rust and Ryan Scales Underachievers | <i>M</i> | −.3 | −.2 | −.2 | −.4 | −.6 | −.8 | .9 |
| | <i>SD</i> | 4.8 | 4.2 | 4.7 | 4.4 | 4.9 | 5.0 | 4.3 |
| Normalachievers | <i>M</i> | −1.6 | .3 | −1.2 | −1.2 | −1.4 | −1.7 | −1.8 |
| | <i>SD</i> | 3.7 | 3.4 | 3.7 | 3.5 | 3.9 | 3.8 | 3.3 |
| Overachievers | <i>M</i> | .3 | .3 | .3 | .2 | 1.0 | .2 | −.4 |
| | <i>SD</i> | 2.9 | 2.8 | 2.8 | 3.4 | 2.7 | 3.0 | 2.8 |
| O minus U | <i>M</i> | .5 | .5 | .4 | .5 | 1.5 | .9 | −1.5 |
| | <i>SD</i> | 6.2 | 5.0 | 5.9 | 5.8 | 5.9 | 6.5 | 5.5 |
| Martin Scales (LAM) AIS-Short | <i>M</i> | 11.8 | 14.7 | 12.5 | 12.7 | 12.2 | 11.4 | 11.8 |
| | <i>SD</i> | 3.0 | 6.9 | 4.5 | 4.1 | 2.5 | 3.3 | 3.0 |
| AIS-Long | <i>M</i> | 43.9 | 41.4 | 43.3 | 43.1 | 45.4 | 42.8 | 43.3 |
| | <i>SD</i> | 6.0 | 6.3 | 6.2 | 6.2 | 5.8 | 5.4 | 6.9 |
| AACH Scale | <i>M</i> | 46.1 | 34.9 | 43.3 | 44.1 | 48.7 | 44.0 | 45.4 |
| | <i>SD</i> | 11.4 | 11.6 | 12.3 | 12.6 | 10.4 | 12.2 | 11.0 |
| Grade Point Average | <i>M</i> | 2.1 | 1.8 | 2.0 | 2.0 | 2.3 | 2.0 | 1.9 |
| | <i>SD</i> | .7 | .6 | .7 | .6 | .7 | .6 | .6 |
| Predicted GPA | <i>M</i> | 2.2 | 2.1 | 2.2 | 2.2 | 2.4 | 2.2 | 1.9 |
| | <i>SD</i> | .2 | .2 | .2 | .2 | .1 | .1 | .2 |

Table 2. It may be noted that most of the students rated themselves well above average in predicting their first semester class standing. The mean scores on the various academic interest scales are roughly comparable to the

mean scores of relevant groups of college students reported in the literature (Campbell & Johansson, 1966; McArthur, 1965).

The intercorrelations of the predictor variables for the A&S students are reported in

TABLE 3
INTERCORRELATIONS OF ACADEMIC INTEREST SCALES FOR ARTS AND SCIENCES STUDENTS

| Scale | U | N | O | O − U | AIS-Short | AIS-Long | AACH | PGPA |
|----------------------|------|--------|--------|--------|-----------|----------|-------|-------|
| Self-rating scale | −.06 | −.03 | .11 | .10 | .05 | .15* | .35** | .22** |
| Rust and Ryan scales | | | | | | | | |
| Underachievers (U) | | −.42** | −.21** | −.88** | −.20** | −.14* | −.09 | −.11 |
| Normalachievers | | | .03 | .34** | .13 | .10 | −.09 | .04 |
| Overachievers (O) | | | | .64** | .42** | .40** | .26** | .16* |
| O minus U | | | | | .35** | .28** | .20** | .18** |
| Martin Scales (LAM) | | | | | | | | |
| AIS-Short | | | | | | .69** | .40** | .06 |
| AIS-Long | | | | | | | .34** | .11 |
| AACH scale | | | | | | | | .13 |

Note.—*n* = 222.
* *p* < .05.
** *p* < .01.

TABLE 4
CORRELATION BETWEEN ACADEMIC INTEREST SCALES AND FIRST SEMESTER
GRADE POINT AVERAGE

| Predictor | Arts and sciences (<i>n</i> = 222) | School of Business Admin. (<i>n</i> = 72) | A & S and SBA (<i>n</i> = 294) | Place- ment (<i>n</i> = 94) | High PGPA (<i>n</i> = 81) | Middle PGPA (<i>n</i> = 85) | Low PGPA (<i>n</i> = 56) |
|----------------------|---|---|--|--|--------------------------------------|--|-------------------------------------|
| Self-rating scale | .00 | .12 | .05 | .19 | -.11 | .00 | -.07 |
| Rust and Ryan Scales | | | | | | | |
| Underachievers (U) | -.18** | -.01 | -.15** | .02 | -.17 | -.13 | -.26 |
| Normalachievers | .08 | .07 | .04 | -.02 | .08 | .03 | .16 |
| Overachievers (O) | .31** | .07 | .26** | .16 | .19 | .31** | .37** |
| O minus U | .30** | .23* | .29** | .06 | .23* | .26* | .40** |
| Martin Scales (LAM) | | | | | | | |
| AIS-Short | .12 | .14 | .06 | .15 | .07 | .16 | .05 |
| AIS-Long | .09 | -.08 | .08 | .10 | .06 | .01 | .10 |
| AACH Scale | .17* | .02 | .18** | .29** | .13 | .14 | .15 |
| Predicted GPA | .19** | .32** | .24** | .20 | .11 | .18 | -.16 |

* $p < .05$.

** $p < 0.1$.

Table 3.⁴ The various measures of academic interest were lowly intercorrelated. None of the r 's for the separate scales (excluding scales which are based in part upon one of the other scales, e.g., the O minus U score or the Martin scales) exceeded .42. Inspection of the item content for the various scales indicated relatively little overlapping (20–40%) in the use of specific items. Surprisingly, 15–25% of the items which did overlap were scored in the opposite direction.

The main findings in the study are reported in Tables 4 and 5. The academic interest

⁴ The intercorrelations of the predictor variables for the other groups of Ss, which are very similar to the intercorrelations for the A & S students, are tabulated in the final research report (Johnson, 1968).

scales predicted first semester performance as effectively as the PGPA regression formula. All of the r 's tended to run fairly low, no r exceeding .40.

The most successful academic interest scales for these students were the Rust and Ryan scales, particularly the Overachiever scale and the O minus U score, and, secondly, the Campbell and Johansson AACH scale. The Martin scales and the self-evaluation scale did not significantly correlate with GPA for any of the groups.

While the direction of the relationship between the SVIB scales and GPA for the three ability groups supported Clark's (1961) finding that the relationship was greater for the lower, or "marginal," students, the r 's were

TABLE 5
MULTIPLE CORRELATIONS BASED ON PREDICTOR VARIABLES WHICH SIGNIFICANTLY ($p < .05$)
INCREASED THE DEGREE OF RELATIONSHIP WITH GRADE POINT AVERAGE

| Group | <i>n</i> | Predictors | <i>R</i> |
|---|----------|--|----------|
| Arts and Sciences | 222 | Overachievers Scale, PGPA | .34 |
| Arts and Sciences | 222 | AACH Scale, PGPA | .24 |
| Arts and Sciences plus School of Business Administration | 294 | O minus U Score, PGPA, Underachievers Scale | .38 |
| Arts and Sciences plus School of Business Administration | 294 | AACH Scale, PGPA | .28 |
| Middle PGPA (Arts and Sciences) | 85 | Overachievers Scale, PGPA | .39 |

not significantly different from each other. A greater number of significant r 's were found with the low and middle groups, however, than for the high group.

The correlations were not any higher for the motivated "placement" group than for the "discussion" groups. Only Campbell and Johansson's scale predicted GPA significantly for this group.

Finally, as shown in Table 5, the magnitude of the relationship between the predicted GPA and GPA was significantly increased by the addition of one or two of the academic interest scales in at least several instances. The total amount of variance accounted for (15 or 16% at most) is still relatively small, but, nonetheless, some of the error in prediction has been reduced.

DISCUSSION

With the exception of Campbell and Johansson's scale, the test-retest reliabilities of the remaining academic interest scales are not sufficiently high for routine individual interpretation. Although the test-retest reliabilities are higher than the reported split-half reliability coefficients (Martin, 1964; Rust & Ryan, 1954), presumably due to the heterogeneous nature of the item content, the reliabilities still average only in the .70 to .75 range. If the scales could be increased in length by using items of comparable validity, the test-retest reliabilities could be substantially improved (Abrahams, 1967). Until such an event, the scales may be most safely used for group interpretations or for forming (not testing) hypotheses regarding individuals.

The modest reliabilities attenuate the maximum validities possible for the scales. Despite this limitation, the Rust and Ryan scales, together with the Campbell and Johansson scale, possessed promising validity for use with the students in this study. The Over-achievers scale, O minus U score, and AACH scale each correlated as highly as predicted GPA with first semester grades.

The ineffectiveness of the Martin scales in predicting GPA needs some explanation. The composition of the student body and/or the courses comprising the first-year schedule apparently varied sufficiently from that of the

University of Pittsburgh to prevent successful cross-validation of the scales. The scales themselves, although constructed in a manner somewhat different from either the Rust and Ryan or Campbell and Johansson scales, do not appear to be at fault in that they did effectively predict academic performance for successive samples at Pittsburgh.

As a one-item measure, the self-evaluation scale may have lacked adequate reliability to predict grade performance. The fact that nearly all the students rated themselves above average suggests that the students' self-perceptions were not very accurate at best. Both Torrance (1954) and Stone (1962) report a similar tendency on the part of students to overestimate their academic potential. Torrance also found very little relationship between self-predicted grades and achieved grades. Stone did not report the predictive validities of the students' self-ratings.

The lack of a significant relationship between self-predicted and obtained grades contrasts sharply with Young's (1954) and O'Hara's (1966) findings that self-ratings added significantly to multiple R s consisting of various aptitude measures in predicting academic success. In both the Young and the O'Hara studies, however, the students made their self-estimates sometime after school had started; in fact, their grades had already been in part determined by examinations which they had taken. It is worth noting that students do respond to feedback they receive within their environment; however, the technique loses any meaning as a preenrollment index of potential academic achievement. The main value of preenrollment ratings, as observed by Torrance (1954), may be to involve the students more deeply in test interpretation and to assist the counselor in determining how resistant the student will be in accepting test results. The scale may reveal the student's wish to succeed but not necessarily reflect any added effort on his part to insure success.

Although clear-cut statistically significant differences among the three predicted GPA groups failed to emerge, the findings were in the expected direction. The difficulty in cross-validating results found with such subdivisions of the total sample has been clearly il-

illustrated by Hakel (1966). Perhaps more carefully refined PGPA groups, as well as a larger n , would have produced more definitive results in the present study. The relationship between ability and achievement at different motivation levels also needs further exploration (French, 1958).

The modified instructions indicating that the test results might be used in advanced placement apparently did not greatly influence the scores for the students. The means and standard deviations for both the "placement" and the "discussion" groups were approximately the same. The predictive validity coefficients were of approximately the same magnitude for each group. If anything, the r 's appeared to run slightly lower for the "placement" group than for the "discussion" group. This result fails to support the hypothesis that "sensible distortion," which may occur in real-life situations (Gellerman, 1963), may actually increase the validity of the test scores. The findings are in accord with Walsh's (1967, 1968) observation that validity of self-report is not greatly affected by incentives to distort. Perhaps more critical instructions (e.g., results would be used in selection instead of advanced placement) or a different test-taking atmosphere (SVIB administered at the same time as entrance examinations) would have had a greater impact on the results.

The success of the Rust and Ryan scales and the AACH scale in contributing significantly to a multiple R based in part upon aptitude and achievement variables in predicting college achievement at the University of Massachusetts is encouraging. The failure of nonintellective measures to cross-validate in predicting academic success in new settings is well recorded (Super & Crites, 1962). Although the size of the multiple R is only moderate, if extreme scores (e.g., plus and minus one standard deviation) are used as cut-offs, relatively accurate classification of successful or unsuccessful students would be possible (Taylor & Russell, 1939). With the collection of local cross-validation data, expectancy tables for converting the scores of very low and very high scoring students into GPA probabilities may be profitably constructed.

Study of the content of the scales, although possibly misleading if not supported by theoretical assumptions, may serve as a source of hypotheses for additional research. The content of the Overachievers scale, perhaps the single most efficient predictor, indicates that achievement beyond one's predicted level is associated with items suggesting conservatism (playing safe, not loaning money), conventionality (lack novel ideas, work where can stay in one place), conscientiousness (plan work in detail), passive feminine interests (birdwatching, music teaching), and lack of mechanical interests (auto mechanic, adjusting a carburetor). According to McArthur (1965), items on the Rust and Ryan scales reflect "conscientious perseverance."

The above description agrees rather well with Nichols' (1966) observation that students who get good grades are likely to be "compulsive and conforming." The achievement level among students may possibly be raised by selecting or training students on such characteristics. Other temperamental or motivational characteristics may be rewarded by modifying the criteria used for achievement within courses or by broadening the definition of achievement to include extracurricular accomplishments. The personal qualities desired on the part of the students will depend upon the criteria of achievement established by the educational institution. Techniques for identifying and reinforcing both the persevering student and the creative student need to be developed.

REFERENCES

- ABRAHAM, N. M. SVIB key length: Dissident data. *Journal of Applied Psychology*, 1967, 51, 266-273.
- CAMPBELL, D. P., & JOHANSSON, C. B. Academic interests, scholastic achievements, and eventual occupations. *Journal of Counseling Psychology*, 1966, 13, 416-424.
- CLARK, K. E. *Vocational interests of nonprofessional men*. Minneapolis: University of Minnesota Press, 1961.
- CLARK, K. E., & CAMPBELL, D. P. *Manual for Minnesota Vocational Interest Inventory*. New York: Psychological Corp., 1965.
- FRENCH, E. G. The interaction of achievement motivation and ability in problem-solving success. *Journal of Abnormal and Social Psychology*, 1958, 57, 306-309.
- GELLMAN, S. W. Personnel testing: What the critics overlook. *Personnel*, 1963, 40(3), 18-26.

- GLOVER, R. H. *Preselection in college admissions at the University of Massachusetts*. Amherst, Mass.: Office of Institutional Studies, University of Massachusetts, 1963.
- HAKEL, M. D. Prediction of college achievement from the Edwards Personal Preference Schedule using intellectual ability as a moderator. *Journal of Applied Psychology*, 1966, 50, 336-340.
- HATHAWAY, S. K. Foreward. In W. G. Dahlstrom & G. S. Welsh (Eds.), *An MMPI handbook: A guide to use in clinical practice and research*. Minneapolis: University of Minnesota Press, 1960.
- HOLLAND, J. L., & LUTZ, S. W. The predictive value of a student's choice of vocation. *Personnel & Guidance Journal*, 1968, 46, 428-434.
- JOHNSON, R. W. The effectiveness of academic interest scales in predicting college achievement. Final Report for United States Office of Education, Contract No. OEC-1-7-068710-3112, May, 1968.
- MARTIN, A. M. The development and successive refinement of an academic interest scale for the Strong Vocational Interest Blank. *Educational and Psychological Measurement*, 1964, 24, 841-852.
- MCARTHUR, C. The validity of the Yale Strong scales at Harvard. *Journal of Counseling Psychology*, 1965, 12, 35-38.
- MCNEMAR, Q. *Psychological statistics*. (3rd ed.) New York: Wiley, 1962.
- NICHOLS, R. C. Nonintellective predictors of achievement in college. *Educational and Psychological Measurement*, 1966, 26, 899-915.
- O'HARA, R. P. Vocational self-concepts and high school achievement. *Vocational Guidance Quarterly*, 1966, 15, 106-112.
- RUCH, F. L., & RUCH, W. W. The K factor as a (validity) suppressor variable in predicting success in selling. *Journal of Applied Psychology*, 1967, 51, 201-204.
- RUST, R. M., & RYAN, F. J. The Strong Vocational Interest Blank and college achievement. *Journal of Applied Psychology*, 1954, 38, 341-345.
- STONE, L. A. Differences between student predictions and statistical predictions of future academic success: The relationship of these differences to scholastic ability. *Studies in Student Personnel Work*, No. 21. Manhattan, Kansas: Student Counseling Center, Kansas State University, September, 1962.
- STRONG, E. K., JR., & CAMPBELL, D. P. *Manual for Strong Vocational Interest Blanks*. Stanford, Calif.: Stanford University Press, 1966.
- SUPER, D. E., & CRITES, J. O. *Appraising vocational fitness*. (Rev. ed.) New York: Harper & Row, 1962.
- TAYLOR, H. G., & RUSSELL, J. T. The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 1939, 23, 565-578.
- TORRANCE, E. P. Some practical uses of a knowledge of self-concepts in counseling and guidance. *Educational and Psychological Measurement*, 1954, 14, 120-127.
- WALSH, W. B. Validity of self-report. *Journal of Counseling Psychology*, 1967, 14, 18-23.
- WALSH, W. B. Validity of self-report: Another look. *Journal of Counseling Psychology*, 1968, 15, 180-186.
- YOUNG, F. C. College freshmen judge their own scholastic promise. *Personnel and Guidance Journal*, 1954, 32, 399-403.

(Received July 8, 1968)

RATED ACCEPTABILITY OF MINERAL TASTE IN WATER: II. COMBINATORIAL EFFECTS OF IONS ON QUALITY AND ACTION TENDENCY RATINGS¹

WILLIAM H. BRUVOLD²

AND

WILLIAM R. GAFFEY

University of California, Berkeley

California State Department of Public Health

Previous research has shown that equal concentration simple solutions of minerals in water receive significantly different taste quality ratings. The differences in ratings appear to be accounted for by solution anions. The present research was designed to investigate the combinatorial effect of mineral anions on taste quality ratings given specially prepared water samples. Thirteen water samples were rated on two scales by 25 Ss. Results indicated that anions independently influence the general taste quality of mineralized water. The implications of these results for establishing limiting standards for minerals in domestic water were discussed.

Research is in progress which aims to provide the information needed to establish limiting standards for mineral content in domestic water in order to ensure potability for daily consumption (Bruvold, Ongerth, & Dillehay, 1967). The current recommendation limiting mineral content made by the United States Public Health Service (1962) is in terms of total dissolved solids (TDS) as milligrams per liter (mg/l) as are the recommendations of the state of California. Standards for mineral content in domestic water expressed only as TDS may be reasonable at present; however, they may not remain so as evidence regarding the relationship between dissolved minerals and taste quality accumulates.

Mineral content in water is composed primarily of calcium, magnesium, potassium, and sodium cations in combination with bicarbonate, carbonate, chloride, nitrate, and sulfate anions. Previous research (Bruvold, 1968; Bruvold & Pangborn, 1966) has amply demonstrated that mean taste quality ratings vary

significantly for the equal concentration mineral solutions containing one cation and one anion. The differences between these ratings were attributed to anion effects since all bicarbonate and sulfate solutions received mildly unfavorable ratings, all chloride solutions moderately unfavorable ratings, and all carbonate solutions strongly unfavorable ratings. The magnitude of differences between ratings given these equal concentration solutions suggests that TDS standards may be inappropriate.

The present research was designed to investigate how taste quality ratings are influenced by various combinations of mineral ions since natural waters used for domestic consumption contain certain amounts of most, if not all, of the ions listed in solution. Such data can answer basic questions regarding the combinatorial effects of ions on taste quality ratings, and they can also be employed in a consideration of various methods for establishing standards which limit mineral content in domestic water.

METHOD

Twenty-five employees of the California State Department of Public Health served voluntarily as raters in this research. Sixteen of the raters were males, 9 were females, and all had served in previous water taste research. All individuals were selected because of their availability and willingness to participate; none was asked to participate on the basis of prior performance in rating the taste of water samples.

¹ This research was supported in part by funds provided by the United States Department of Interior, Office of Water Resources Research, as authorized under the Water Resources Act of 1964, by the University of California Water Resources Center, and in part by United States Public Health Service, National Institutes of Health General Research Support Award 5-S01-FR-5441 to the School of Public Health, University of California, Berkeley.

² Requests for reprints should be sent to William H. Bruvold, University of California, School of Public Health, Earl Warren Hall, Berkeley, California 94720.

TABLE 1
CHEMICAL ANALYSIS DATA FOR 13 WATER SAMPLES

| Sample number | Na | Cl | CO ₃ | HCO ₃ | SO ₄ | TDS by evaporation |
|---------------|-----|-----|-----------------|------------------|-----------------|--------------------|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 176 | 125 | 31 | 195 | 0 | 432 |
| 3 | 157 | 0 | 38 | 195 | 124 | 424 |
| 4 | 161 | 0 | 42 | 348 | 0 | 374 |
| 5 | 173 | 125 | 0 | 100 | 124 | 474 |
| 6 | 428 | 375 | 132 | 305 | 0 | 1119 |
| 7 | 400 | 0 | 138 | 348 | 377 | 1114 |
| 8 | 422 | 0 | 312 | 525 | 0 | 1005 |
| 9 | 400 | 380 | 0 | 0 | 376 | 1188 |
| 10 | 888 | 760 | 300 | 572 | 0 | 2212 |
| 11 | 786 | 0 | 264 | 680 | 751 | 2180 |
| 12 | 742 | 0 | 288 | 1464 | 0 | 1732 |
| 13 | 841 | 760 | 0 | 0 | 750 | 2380 |

Note.—Concentrations in milligrams per liter (mg/l).

Twelve water samples were specially prepared using reagent grade NaCl, NaHCO₃, and Na₂SO₄ with double distilled water. Sample concentrations were designed to cover the meaningful TDS range found in natural waters used for domestic supply, and to vary widely in type and amount of anions present. Double distilled water was also employed as one sample in the study. Chemical analysis data for all water samples obtained 2 wk. after preparation are shown in Table 1. Sodium was the only cation used since two earlier attempts to prepare samples containing calcium and magnesium cations resulted in uncontrolled carbonate precipitation. None of the 12 samples showed any sign of precipitation over the entire duration of the study.

TABLE 2
MEANS AND STANDARD DEVIATIONS
FOR 13 WATER SAMPLES

| Sample number | Quality ratings | | Action tendency ratings | |
|---------------|-----------------|------|-------------------------|------|
| | M | SD | M | SD |
| 1 | 7.27 | 1.57 | 7.41 | 1.37 |
| 2 | 6.32 | 1.32 | 6.80 | 1.48 |
| 3 | 6.40 | 1.51 | 6.90 | 1.51 |
| 4 | 5.95 | 1.45 | 6.50 | 1.27 |
| 5 | 6.44 | 1.34 | 6.81 | 1.40 |
| 6 | 3.95 | 1.32 | 4.50 | 1.59 |
| 7 | 4.58 | 1.13 | 5.24 | 1.28 |
| 8 | 3.76 | 1.37 | 4.32 | 1.50 |
| 9 | 4.78 | 1.21 | 5.55 | 1.88 |
| 10 | 2.68 | 0.94 | 3.22 | 1.35 |
| 11 | 3.45 | 1.32 | 4.00 | 1.37 |
| 12 | 3.78 | 1.17 | 4.42 | 1.57 |
| 13 | 3.35 | 1.03 | 3.76 | 1.60 |

A quality and an action tendency rating scale were employed in this study. Construction of these scales has been fully reported in an earlier paper (Bruvold, 1968). Both scales contained nine rating statements and each statement referred directly to the taste of water. Scale distance between individual items was approximately equal for both sets of statements and scale values were highest for statements indicating the most favorable reaction to the water's taste.

Each rater took part in three rating sessions separated by a period of 1 wk. All 13 samples were rated during each session. Order of sample presentation was independently randomized before each session, and a different sample code was employed for each of the 3 wk. of the study. Rating was performed alone in a small air-conditioned room whose temperature was maintained at 72° ± 2° F. Samples were served at room temperature in 100 ml. beakers filled to the 75 ml. level. Each rater tasted each sample three times, marked the appropriate rating category for the quality and the action tendency scale, rinsed thoroughly with Berkeley tap water (85 mg/l TDS), and rested for 30 sec. as gauged by a laboratory timer. This procedure was repeated until all samples were rated. Numbers were not used in conjunction with the rating scales; raters simply marked the appropriate category for each sample code and rating statement shown on the appropriate data sheet. Quality statements were ordered with the most favorable item at the top of the data sheet, while action tendency items were listed in reverse order placing the most unfavorable statement at the top of the data sheet.

RESULTS

The 3 ratings given each water sample by a rater on each scale were averaged yielding 13 mean ratings for each scale and rater. Equal appearing interval scale values for rat-

TABLE 3
FREQUENCY DISTRIBUTIONS FOR MULTIPLE REGRESSION COEFFICIENTS AND TERMS
FROM THE INDIVIDUAL REGRESSION EQUATIONS

| Multiple <i>R</i> | Frequency | | Regression equation constants | Frequency | |
|-------------------|-----------|----|-------------------------------|-----------|----|
| | Q | AT | | Q | AT |
| 0.95-0.99 | 5 | 6 | 9.50-10.99 | 1 | 0 |
| 0.90-0.94 | 11 | 11 | 8.00-9.49 | 4 | 6 |
| 0.85-0.89 | 5 | 6 | 6.50-7.99 | 9 | 13 |
| 0.80-0.84 | 4 | 1 | 5.00-6.49 | 10 | 4 |
| 0.75-0.79 | 0 | 1 | 3.50-4.99 | 1 | 2 |

| Regression equation terms | Na | | HCO ₃ | | CO ₃ | | Cl | | SO ₄ | |
|---------------------------|----|----|------------------|----|-----------------|----|----|----|-----------------|----|
| | Q | AT | Q | AT | Q | AT | Q | AT | Q | AT |
| .060-.089 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| .030-.059 | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| .000-.029 | 14 | 13 | 6 | 8 | 4 | 6 | 4 | 6 | 4 | 6 |
| -.030-.001 | 4 | 5 | 19 | 16 | 12 | 15 | 17 | 16 | 19 | 18 |
| -.060-.031 | 0 | 2 | 0 | 1 | 9 | 4 | 4 | 3 | 2 | 1 |

ing statements were used as individual numerical rating scores. Means and standard deviations for ratings given the 13 solutions are shown in Table 2. Each entry in Table 2 is based upon 25 mean rating scores.

Stepwise multiple regression analyses (Walker & Lev, 1953) were performed on the mean ratings for each individual rater and scale. The results of these 50 analyses, analyses which employed the scale rating as the dependent variable and the ionic concentration

values shown in Table 1 as independent variables, are summarized in Table 3. The same regression analysis was performed upon all 325 mean ratings for each scale. The summary regression equation for the quality scale was $Q' = 6.81 + 0.020 (\text{Na}) - 0.008 (\text{HCO}_3) - 0.022 (\text{CO}_3) - 0.016 (\text{Cl}) - 0.011 (\text{SO}_4)$, the multiple *R* associated with this equation was 0.719, and eta was 0.738. Analogous values for the action tendency scale were $AT' = 7.22 + 0.011 (\text{Na}) - 0.005$

TABLE 4
ANALYSIS OF VARIANCE FOR RATING SCALE RESULTS BY TDS CATEGORIES

| Source | df | Quality ratings | | | Action tendency ratings | | |
|-----------------|-----|-----------------|--------|----------|-------------------------|--------|----------|
| | | SS | MS | F | SS | MS | F |
| Between people | 24 | 311.08 | | | 458.89 | | |
| Within people | 300 | 889.20 | | | 851.60 | | |
| Waters | 12 | 655.33 | 54.61 | 67.42** | 594.61 | 49.55 | 55.67** |
| Linear trend | 1 | 566.61 | 566.61 | 699.52** | 512.17 | 512.17 | 575.47** |
| Quadratic trend | 1 | 43.79 | 43.79 | 54.06** | 20.48 | 20.48 | 23.01** |
| Cubic trend | 1 | 0.55 | 0.55 | 0.68 | 0.81 | 0.81 | 0.91 |
| Quartic trend | 1 | 2.79 | 2.79 | 3.44 | 4.59 | 4.59 | 5.16* |
| Residual | 288 | 233.87 | 0.81 | | 256.99 | 0.89 | |
| Total | 324 | 1200.28 | | | 1310.49 | | |

* $p < .05$.

** $p < .001$.

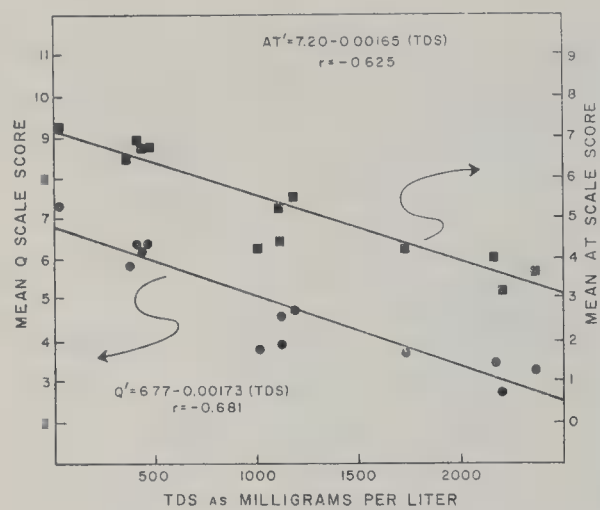


FIG. 1. Lines of best fit for Q and AT scale ratings.

$(\text{HCO}_3) - 0.016$ $(\text{CO}_3) - 0.010$ $(\text{Cl}) - 0.007$ (SO_4) , $R = 0.659$, and $\text{eta} = 0.675$.

Regression analyses of the relationship between TDS values and mean scale ratings were also performed (Winer, 1962). Results of these analyses are shown in Table 4 and in Figure 1. Quadratic regression equations for the TDS values were $Q' = 7.50 - 0.0036$ $(\text{TDS}) + 0.0000008$ $(\text{TDS})^2$ and $\text{AT}' = 7.71 - 0.0030$ $(\text{TDS}) + 0.0000005$ $(\text{TDS})^2$.

DISCUSSION

The multiple regression equations reported above involve only one cation, sodium, since earlier attempts to prepare solutions containing additional common mineral cations resulted, as noted, in uncontrolled carbonate precipitation. The relatively large regression coefficient for sodium was likely due to the fact that there was a substantial direct relationship between amount of sodium and TDS in the solutions here employed. The regression coefficients for anions match earlier re-

sults for simple solutions (Bruvold, 1968; Bruvold & Pangborn, 1966) since the negative coefficients are smallest for bicarbonate and sulfate, intermediate for chloride, and largest for carbonate.

Multiple regression analyses demonstrated that combinatorial ionic effects on quality and action tendency ratings were adequately described by a model involving only a first degree function containing no interaction terms. The multiple correlation coefficients for the two scales were close enough, in light of the degrees of freedom involved, to the corresponding etas to show that adding higher degree or interaction terms to the five-variable function could account for only a very small additional portion of the explainable variance. These data indicate that, for the raters, ions, and concentrations here employed, there were no important synergistic or masking effects between ions as they influenced taste ratings. Rather, each ion appeared to make a straightforward contribution to the ratings obtained according to its concentration in solution without affecting or being affected by other ionic constituents. The same interpretation holds for individual results since multiple correlation coefficients were generally high as shown in Table 3. Further, most individual regression equations were very similar to the summary regression equations reported.

A further opportunity to test the validity of independent ionic influence on taste quality ratings arises in connection with simple mineral solutions rated a year before the present study was undertaken (Bruvold, 1968). If the conclusion of independence is valid, it should be possible to reproduce the earlier ratings using only the multiple regression equations here derived. Following this reasoning, pre-

TABLE 5
PREDICTED AND OBSERVED MEAN RATINGS FOR SIMPLE SOLUTIONS

| Mineral | Concentration | Mean Q rating | | Mean AT rating | |
|---------------------------------|---------------|---------------|----------|----------------|----------|
| | | Predicted | Observed | Predicted | Observed |
| NaHCO ₃ | 1,000 mg/l | 6.48 | 5.15 | 6.60 | 5.79 |
| Na ₂ SO ₄ | 1,000 mg/l | 5.85 | 5.79 | 6.05 | 6.88 |
| NaCl | 1,000 mg/l | 5.10 | 4.49 | 5.56 | 5.43 |
| Na ₂ CO ₃ | 1,000 mg/l | 3.04 | 3.11 | 2.93 | 3.58 |

dicted mean Q and AT ratings were obtained for each of four simple mineral solutions by introducing appropriate values for the anion and cation in question, setting remaining ionic values equal to zero, and solving the regression equation in a straightforward fashion. Predicted mean ratings obtained in this manner are shown in Table 5 together with the mean ratings actually obtained from the earlier study. It may be seen that the agreement between predicted and observed mean ratings was good considering variations in procedures and raters (11 raters were common to both studies). Such agreement would not have been obtained if ions combine in a complex manner to determine taste ratings.

Results of the regression analyses involving TDS, rather than individual ions, showed that there was a significant linear trend in the data which accounted for the bulk of the explainable variance. Thus, multiple regression was not markedly superior to simple regression in terms of total variance accounted for. However, inspection of Figure 1 shows that certain means deviated considerably from the line of best fit. This result indicates that the multiple regression approach is preferable to the simpler TDS approach for establishing limiting standards since it would not mispredict rating scores as grossly as the simpler approach for unusual water samples containing very high concentrations of some ions and very low concentrations of others. Such mispredictions could cause great difficulty in attempts to establish and administer limiting standards for mineral content in domestic water. TDS standards could result in approval of unpotable waters for daily consumption while requiring demineralization for waters actually suitable for daily drinking.

The statistically significant, but not highly important, quadratic trends found in the TDS regression analyses were probably due to the fact that mean carbonate concentrations, concentrations impossible to control precisely

during solution preparation, were relatively higher for the 1,100 mg/l samples than they were for the remaining TDS levels. Had the relative carbonate concentrations been lower for the 1,100 mg/l samples, a negative quadratic term might have been obtained. Thus, the multiple regression approach also appears preferable to second or third degree TDS functions for establishing limiting standards. The present results suggest that the nature of these higher order terms was related to the relative pattern of ionic concentrations in the samples studied. Different patterns of such concentrations would require separate equations to describe adequately each set of data, while, conceivably, one multiple linear regression equation could adequately describe all such data sets.

Therefore it is concluded on the basis of the present results that a multiple linear regression equation using separate ionic concentrations as independent variables should provide the best method for establishing limiting standards for mineral content in domestic water. Consumer ratings of natural waters (Bruvold et al., 1967) will provide the data for establishing standards and for further evaluation of conclusions regarding independent anion effects based on this research.

REFERENCES

- BRUVOLD, W. H. Scales for rating the taste of water. *Journal of Applied Psychology*, 1968, **52**, 245-253.
- BRUVOLD, W. H., ONGERTH, H. J., & DILLEHAY, R. C. Consumer attitudes toward mineral taste in water. *Journal American Water Works Association*, 1967, **59**, 547-556.
- BRUVOLD, W. H., & PANGBORN, R. M. Rated acceptability of mineral taste in water. *Journal of Applied Psychology*, 1966, **50**, 22-32.
- Drinking water standards*. Washington, D. C.: Public Health Service, United States Department of Health, Education and Welfare, 1962.
- WALKER, H. M., & LEV, J. *Statistical inference*. New York: Holt, 1953.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received July 18, 1968)

PERSONALITY INVENTORY CORRELATES OF CREATIVITY AMONG ARCHITECTS¹

WALLACE B. HALL AND DONALD W. MACKINNON²

*Institute of Personality Assessment and Research
University of California, Berkeley*

Multiple stepwise correlation analyses were conducted on each of seven personality inventories against the rated creativity of 62 architects from a nationwide sample. In each analysis the best combination of three variables was identified. The equations derived from the analyses were then validated on a second sample of 62 architects. Initial multiple correlations ranged from .80 to .35; cross-validated coefficients ranged from .55 to .20. Specific cross-validated coefficients were as follows: Adjective Check List (ACL), .38; California Psychological Inventory (CPI), .47; FIRO-B, .41; Minnesota Multiphasic Personality Inventory (MMPI), .20; Myers-Briggs Type Indicator (MBTI), .42; Strong Vocational Interest Blank (SVIB), .55; and Allport-Vernon-Lindzey Study of Values (A-V-L), .38.

An assessment study of a nationwide sample of American architects (MacKinnon, 1962, 1965) provided an opportunity to examine the relative validity of a variety of personality inventories, using ratings of creativity as a criterion.

Architects were chosen for study on the grounds that they should clearly manifest those personality characteristics which are typical of the creative person. Behind this reasoning lies the observation that if an architect's designs are to give delight, the architect must be an artist, and if they are to be technologically sound and efficiently planned, he must also be something of a scientist, at least an applied scientist or engineer. Yet surely it is not sufficient that an architect be both artist and scientist if he is to be highly successful in the practice of his profession. He must also to some extent be businessman, lawyer, advertiser, author-journalist, psychiatrist, educator, and psychologist.

The total sample consisted of 124 American architects. Forty of them, constituting a nationwide sample and here designated as Architects 1, were nominated by a panel of five professors of architecture at the University of

California, Berkeley, for the unusual creativeness they had shown in the practice of their profession.

The second group, Architects 2, consisted of 43 architects chosen so as to match Architects 1 with respect to age and the geographic location of their practice. Each of them met the additional requirement that he had had at least 2 yr. of work experience and association with one of the originally nominated creative architects.

The third sample, Architects 3, was also chosen to match Architects 1 with respect to age and geographic location of practice, but, unlike Architects 2, the 41 men in this group had never worked with any of the Architects 1.

The three samples were selected in this manner in the hope of tapping a range of creative talent sufficiently wide to be fairly representative of the profession as a whole. To determine whether or not this objective was met, ratings on a 7-point scale of the creativity of all 124 architects were obtained from six groups of architects and architectural experts: the 5 members of the original nominating panel at the University of California, 19 professors of architecture distributed nationwide, 6 editors of the major American architectural journals, 32 Architects 1, 36 Architects 2, and 28 Architects 3.

The mean intercorrelation among these six rating groups provides an estimate of the reliability of the ratings, namely, .84. The mean

¹ This is a slightly modified version of a paper presented at the annual meeting of the Western Psychological Association, Honolulu, Hawaii, June 14-19, 1965.

² Requests for reprints should be sent to Donald W. MacKinnon, Director, Institute of Personality Assessment and Research, University of California, 2240 Piedmont Avenue, Berkeley, California 94720.

ratings of creativity for the three groups are Architects 1, 5.46; Architects 2, 4.25; and Architects 3, 3.54. The differences are in the expected direction and are statistically significant ($p < .001$). In other words, the three groups do represent significantly different levels of creativeness. At the same time, it must be noted that the three samples show an overlap in their judged creativeness; that is, they are not discrete and discontinuous. When the three groups are combined, the ratings approximate a normal distribution of judged creativeness ranging from a low of 1.9 to a high of 6.5 on a 7-point rating scale.

Of the seven inventories to be examined, the SVIB (Strong, 1959) clearly yielded the largest number of significant correlations with the criterion. Of the 57 scales which were scored, 40 correlated significantly ($p < .05$), ranging from Artist, with a coefficient of .59, to Banker, with a coefficient of $-.66$.³ The 5 scales showing the highest positive correlation with the criterion were Artist, .59; Author-journalist, .54; Lawyer, .44; Advertising man, .42; Musician, .38. Those with the highest negative correlations were Banker, $-.66$; Office man, $-.60$; Accountant, $-.54$; Policeman, $-.52$; and Purchasing Agent, $-.50$.

In addition to this consideration of individual scales, all 57 scales of the SVIB were used to derive an optimum three-variable multiple regression equation, using as a criterion the architects' rated creativity.

In order that the multiple regression equation might be cross-validated, the total sample of 124 architects was divided into two subsamples of 62, matched on creativity. From the roster of Architects 1, 2, and 3, listed according to the rank order of their creativity within each sample, odd-numbered Ss were assigned to the sample on which the

TABLE 1
RELATIONSHIP OF SEVEN PSYCHOLOGICAL TESTS TO
RATINGS OF CREATIVITY, IN TWO
SAMPLES OF ARCHITECTS

| Test | Regression coefficients | R (Initial sample ^a) | r (Cross-validating sample ^b) |
|-----------------------------|-------------------------|----------------------------------|---|
| SVIB (57 variables) | | | |
| Office man | -.543 | | |
| Banker | -.503 | .802* | .551* |
| President-Mfg. Concern | -.258 | | |
| CPI (18 variables) | | | |
| Sp | .547 | | |
| Ac | -1.015 | .572* | .471* |
| Fe | .990 | | |
| MBTI | | | |
| E (extraversion) | -.350 | | |
| N (intuition) | .386 | .534* | .417* |
| P (perception) | .556 | | |
| FIRO-B (6 variables) | | | |
| E ¹ | -1.847 | | |
| E ^c | .970 | .631* | .406* |
| W ^c | -1.077 | | |
| A-V-L (6 variables) | | | |
| Economic | -.786 | | |
| Social | -.272 | .640* | .382* |
| Religious | -.181 | | |
| ACL (24 variables) | | | |
| Self-confidence (Gough key) | -.302 | | |
| Autonomy (Heilbrun key) | .212 | .607* | .376* |
| Change (Heilbrun key) | .413 | | |
| MMPI (13 variables) | | | |
| L | .980 | | |
| F | .808 | .348* | .197 |
| Pa (with K correction) | .718 | | |

^a $n = 62$.

^b $n = 62$.

* $p < .01$.

multiple regression solution was derived, even-numbered Ss to the sample on which the equation was cross-validated.

The computational program which was employed proceeds by steps, each of which yields a trial set of predictors and corresponding regression statistics. At each step, one variable is added or deleted to produce the next trial set. The choice of variable guarantees either a larger set which is significantly better, or a smaller set which is not significantly worse. Under this procedure the new

³ A 2-page table giving correlations with rated creativity for the total sample of 124 architects of all variables scored on the 7 personality inventories examined has been deposited with the National Auxiliary Publications Service. Order Document No. 00458 from National Auxiliary Publications Research and Microfilm Publication, Inc. Remit in advance \$3.00 for photocopies or \$1.00 for microfiche and make checks payable to: Service of the American Society for Information Service, Inc., 22 West 34th Street, New York, New York 10001.

set is the best that can be obtained by a single addition or deletion. In order to insure comparability of results across tests, an arbitrary decision was made to stop the analysis after three variables had been selected; thus, for each test, there would be an optimum three-variable equation.⁴

As Table 1 shows, the three SVIB variables selected for the multiple regression, all with negative weights, were the scales for Office Man, Banker, and President-Manufacturing Concern. The multiple regression coefficient between these three scales and the criterion of rated creativity was .80. In the cross-validating sample the correlation dropped to .55, a still highly significant value.

In the case of the CPI (Gough, 1964), 11 of the 18 scales had significant correlations with the architects' rated creativity: Fe (femininity), .24; Fx (flexibility), .24; Sa (self-acceptance), .19; Sp (social presence), .18; Cm (communality), $-.31$; Sc (self-control), $-.31$; Ac (achievement via conformance), $-.24$; Gi (good impression), $-.23$; To (tolerance), $-.21$; Re (responsibility), $-.20$; and Wb (well-being), $-.20$. The three variables selected in the multiple-regression solution were Sp and Fe with positive weights, and Ac with a large negative weight. The initial validity of this CPI equation was .57 which on cross-validation became .47.

Of the eight scales of the MBTI (Myers, 1962) which tests Jungian functions and attitudes, four were significantly correlated with creativity in the total sample of architects: Intuition, .45; Perception, .40; Sensing, $-.41$; and Judgment, $-.29$.

The three variables of the MBTI selected by the IBM program for the multiple regression equation were Extraversion with a negative weight and Intuition and Perception, with positive weights. The equation correlated .53 with the criterion in the first sample, and .42 in cross-validation.

On Schutz's (1967) test of interpersonal behavior, FIRO-B, four of the six scales were significantly correlated with the criterion: E^I

(desire to include others in one's activities) correlates $-.44$, W^I (desire to be included in others activities) correlates $-.26$, W^C (desire to be controlled by others) $-.24$, and E^C (desire to control others) .34.

The three scales selected by the computer-programmed multiple regression solution are E^I and W^C, both with negative weights, and, with a positive weight, E^C. The initial validity of the equation was .63, which on cross-validation dropped to .41, again a still highly significant correlation.

On the A-V-L (Allport, Vernon, & Lindzey, 1960), three variables were significantly correlated with the criterion in the total sample of 124: the theoretical value, .18; the aesthetic value, .35; and the economic value, $-.48$. (For $n = 124$, a value of .18 is significant at the .05 level, a value of .23 significant at the .01 level.)

However, the three A-V-L variables selected by the IBM program for the multiple regression equation were the economic, social, and religious values, all with negative weights. The initial correlation of the equation was .64 which on cross-validation dropped to .38. The cross-validated value is still significant, since with an n of 62 a value of .25 is significant at the .05 level, a value of .32 significant at the .01 level.

The ACL (Gough & Heilbrun, 1965) may be scored for 24 variables, and scores on 14 of these scales were significantly correlated with the criterion in the total sample of architects. Those correlating positively were Change, .45; Exhibition, .36; Autonomy, .36; Aggression, .34; Lability, .27, and Number of Unfavorable Adjectives Checked, .24; while those showing negative correlations were Self-control, $-.39$; Deference, $-.35$; Personal Adjustment, $-.33$; Order, $-.31$; Nurturance, $-.30$; Intraception, $-.25$; Affiliation, $-.24$; and Endurance, $-.23$.

The three scales of the ACL selected for the multiple regression solution were Self-confidence with negative weighting, and Autonomy and Change both positively weighted. The equation correlated .61 with the criterion in the first sample and .38 in cross-validation.

Of the seven inventories which were administered, the MMPI (Hathaway & McKinley, 1951) fared least well. In the total

⁴ An empirical check with the SVIB, using 4-scale and 5-scale equations, revealed greater shrinkage in cross-validation than for the 3-scale equation. The 3-scale combination, that is, has an additional advantage of greater stability in cross-validation.

TABLE 2

MEANS, STANDARD DEVIATIONS, AND CORRELATIONS OF COMPUTED SCORES FOR CROSS-VALIDATING SAMPLE OF 62 ARCHITECTS

| Test | <i>M</i> | <i>SD</i> | Correlations | | | | | | |
|--------|----------|-----------|--------------|------|------|------|-------|-----|------|
| | | | SVIB | CPI | MBTI | FIRO | A-V-L | ACL | MMPI |
| SVIB | 440.1 | 87.1 | — | | | | | | |
| CPI | 430.4 | 44.9 | .19 | — | | | | | |
| MBTI | 440.9 | 62.9 | .68* | .34* | — | | | | |
| FIRO-B | 428.8 | 56.9 | .52* | .19 | .44* | — | | | |
| A-V-L | 438.6 | 78.0 | .70* | .23 | .56* | .37* | — | | |
| ACL | 452.2 | 48.2 | .54* | .39* | .57* | .37* | .43* | — | |
| MMPI | 447.9 | 36.6 | .15 | .17 | .12 | .20 | .34* | .06 | — |

* $p < .01$.

sample, only 4 of the 13 regular scales had significant correlations with the criterion: Mf (femininity) .29, F (validity) .28, Pd (psychopathic deviate) .23, and Sc (schizophrenia) .20.

The validity of the multiple regression equation for the MMPI was .35 in the first sample, but shrank to .20 ($p > .10$) on cross-validation. The three variables in the equation were L (lie), F (validity), and Pa (Paranoia), all with positive weights.

Table 2 presents means and standard deviations for the seven scores derived from the multiple regression solutions computed for the cross-validation sample of 62 architects together with the score intercorrelations.

Of the seven tests whose validities have been reviewed, it is interesting to note that it is the SVIB, not always seen as a personality measure, which surpassed the others in its ability to forecast the rated creativity of the architects.

It is likewise worth noting that the MMPI, which would be thought of by many to be an instrument of choice in the study of creativity because of its relevance to psychopathology and ego dysfunctions having motivational implications, is relatively weak.

A problem with regression equations containing a large number of variables is that it is difficult to interpret them psychologically. In the present instance, using a three-variable solution, one is struck by the good psychological sense that can be made of the several equations. For example, the CPI equation,

+ .547 Sp - 1.015 Ac + .990 Fe, emphasizes factors found repeatedly to characterize more creative persons. The spontaneity and self-confidence reflected in the Social Presence scale belies one stereotype of creativity, that of the socially anxious and ineffective misfit. A large negative weight given to Achievement via Conformance goes along with the frequently found lower score on this scale relative to Achievement via Independence, which is a combination quite characteristic of both highly creative and successful persons in the professions. The Femininity scale weighting reflects an often repeated finding that our creative Ss reveal an openness to their own feelings and emotions, a sensitive intellect and understanding self-awareness, with widely ranging interests including many which in Western culture are thought of as feminine.

Another example may be found in the MBTI equation, - .350 E + .386 N + .556 P. The creative architect tends to be less often extraverted in the Jungian sense. To oversimplify the case, the intense personal and social interaction required with many clients is not to his liking, and he would prefer time for contemplative thought and creative activity. However, as noted above for the CPI, one should not lose sight of the fact that he does interact with others with marked social presence, often with consummate skill. The creative architect prefers, as well, intuitive perception to the more prosaic and commonly found direct sensing and controlled, planned, and orderly judgmental approach to all ex-

perience. There is a decided preference for openness and receptivity, both to experience and new ideas, as well as for a concern with deeper meanings and possibilities inherent in things and situations.

This paper has focused solely on concurrent validities; one must recognize that a more important problem is whether or not, and to what extent, these personality inventories would demonstrate predictive validities (over time) of a comparable magnitude. This is an empirical question which only future research can answer, but the findings of this cross-sectional inquiry give some reason to hope that personality inventories may be proved to possess longitudinal as well as concurrent utility.

REFERENCES

- ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *Study of Values: Manual of directions*. (Rev. ed.) Boston: Houghton Mifflin, 1960.
- GOUGH, H. G. *Manual for the California Psychological Inventory*. (Rev. ed.) Palo Alto, California: Consulting Psychologists Press, 1964.
- GOUGH, H. G., & HEILBRUN, A. B., JR. *The Adjective Check List manual*. Palo Alto, California: Consulting Psychologists Press, 1965.
- HATHAWAY, S. R., & MCKINLEY, J. C. *Minnesota Multiphasic Personality Inventory manual*. (Rev. ed.) New York: Psychological Corporation, 1951.
- MACKINNON, D. W. The nature and nurture of creative talent. *American Psychologist*, 1962, 17, 484-495.
- MACKINNON, D. W. Personality and the realization of creative potential. *American Psychologist*, 1965, 20, 273-281.
- MYERS, I. B. *Manual (1962): The Myers-Briggs Type Indicator*. Princeton, New Jersey: Educational Testing Service, 1962.
- SCHUTZ, W. C. *The FIRO scales: Manual*. Palo Alto, California: Consulting Psychologists Press, 1967.
- STRONG, E. K., JR. *Manual for Strong Vocational Interest Blanks for Men and Women*. (Revised blanks; Forms M and W). Palo Alto, California: Consulting Psychologists Press, 1959.

(Received July 18, 1968)

VALIDITY, PREDICTIVE EFFICIENCY, AND PRACTICAL SIGNIFICANCE OF SELECTION TESTS

ERVIN W. CURTIS¹ AND EDWARD F. ALF

*United States Naval Personnel Research Activity, San Diego*²

Despite many warnings, validity coefficients continue to be accepted at face value as measures of practical significance. This practice is herein evaluated by examining each functional relationship between three indexes of predictive efficiency— r , r^2 , and E —and three measures of practical significance—the increase of the criterion mean, the expected proportion “satisfactory,” and the expected proportion in 10 criterion categories. The validity coefficient, r , is a linear function of the increase of the criterion mean and very nearly a linear function of the other two measures of practical significance; r^2 and E are related to these three measures in a more curvilinear manner. A table is presented that gives the proportion expected in each of 18 criterion categories as a function of r and the selection ratio.

Brogden (1946) demonstrated that a selection test's validity, r , is a linear function of the difference between two criterion means: the mean for the group above the predictor cutoff and the mean for the population. Even more important, he showed that r equals the proportion improvement over chance that is possible with each selection ratio. Although these facts are not mentioned in most of the textbooks in the area of personnel selection (Dunnette, 1966; Ghiselli & Brown, 1948; Guilford, 1954, 1965; Guion, 1965; Horst, 1966; Nunnally, 1959; Thorndike, 1949), their importance is underscored by the pervasive tendency of psychologists to consider the validity coefficient not just as a measure of correlation, but as a measure of practical significance. Thus it is common practice to consider the correlation between a selection test and a criterion as an indication of the value of the test for the institution using the test. Brogden's analysis shows that this is a reasonable thing to do, provided that the increase of the criterion mean is most important to the institution. An unanswered question is: How does r relate to other measures of practical significance?

The facts that Brogden demonstrated also bear on another question that has not been adequately discussed in the psychological

literature: Which index of predictive efficiency, r , r^2 , or E ($E = 1 - \sqrt{1 - r^2}$), is the best index of practical significance? If the increase of the criterion mean is a linear function of r , it cannot be a linear function of r^2 or E . Therefore, in terms of this measure of practical significance, r is the better index. However, there are other measures of practical significance that are functions of r , r^2 , and E . Which index of predictive efficiency is the best index of practical significance for each measure of practical significance?

This paper explores the functional relationship between each index of predictive efficiency and each of three measures of practical significance: (a) the increase of the criterion mean, (b) proportion satisfactory from the Taylor-Russell tables (1939), and (c) proportions in 10 criterion categories.

INCREASE OF THE CRITERION MEAN

Brogden (1946) related r to the increase of the criterion mean due to the selection test. He used the standard-score formula for r to show that

$$r = \frac{Z_S}{Z_H} = \frac{M_S - M_P}{M_H - M_P} \quad [1]$$

where

M_S is the criterion mean of the selected group
 M_P is the criterion mean of the population
 M_H is the criterion mean of the upper tail of the criterion distribution equal in number to the selected group

¹ Requests for reprints should be sent to Ervin W. Curtis, 8853 Alpine, La Mesa, California 92041.

² The contents of this paper do not necessarily represent the official position or policy of the Department of the Navy.

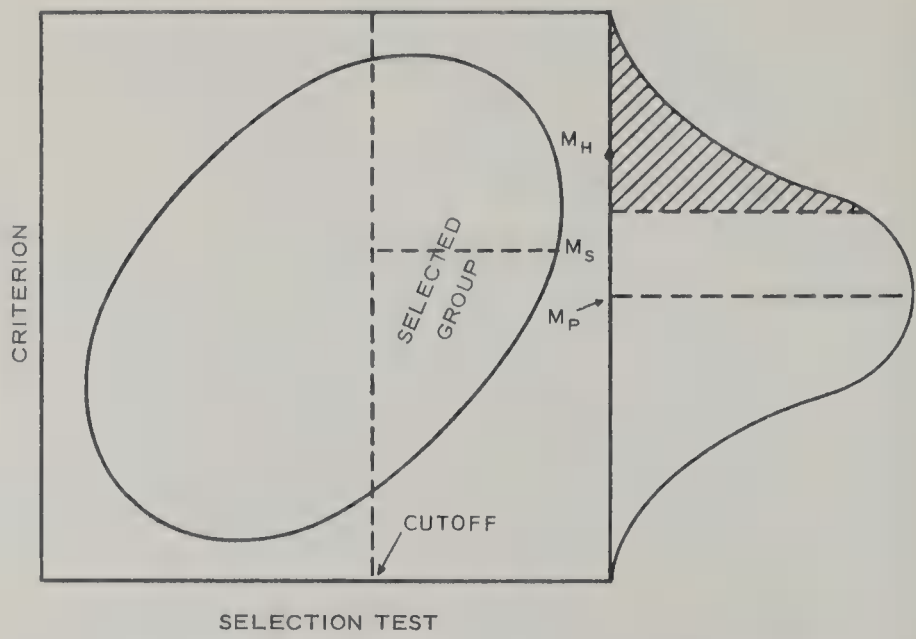


FIG. 1. Correlation surface and criterion distribution showing the means of the “high” group, the selected group, and the population.

Z_S is the z score of M_S
 Z_H is the z score of M_H .

These means are shown in Figure 1. The unfamiliar one, M_H , is the mean of the selected group when the selection test is a perfect predictor of the criterion, that is, when r equals 1.00. This mean provides the upper limit for M_S , the mean of the selected group. The lower limit is M_P if negative correlation is ruled out.

The numerator in Formula 1, $M_S - M_P$, is the actual increase of the criterion mean, while the denominator is the largest possible increase. Therefore, r equals the proportion of possible increase actually achieved. For example, an r of .05 indicates that the test provides 5% of the improvement over chance that a perfect test would provide; an r of .50, 50%; an r of .95, 95%; etc.

Cross-multiplying Formula 1 yields

$$r(M_H - M_P) = M_S - M_P. \tag{2}$$

Since $M_H - M_P$ is a constant when the proportion selected is held constant, $M_S - M_P$ is a linear function of r . Figure 2 shows this linear relationship. The curves for r^2 and E are shown for comparison.

Thus, Brogden showed that r is a linear function of the difference between the crite-

ri-
on
means of the selected group and the population. This implies that the units of the r scale have equal value for the institution using the test, which is in sharp contrast to the implications of r^2 and E . They imply that the units at the high end of the r scale are much more important than the units at the low end. For example, a little computation shows that E increases 30 times as much when r increases from .90 to .95 as when r increases from .05 to .10. In contrast, Brogden showed that the criterion mean increases equally in the two cases.

PROPORTION SATISFACTORY

The mean rise in criterion scores is not the only important effect of selection tests upon criterion distributions. Another effect is the change in the proportion of selectees above a critical point on the criterion—a point that has special significance for the institution using the test. Taylor and Russell (1939) provided tables that give the proportion of selectees expected to be “satisfactory” on the criterion for any combination of r , base rate, and selection ratio.

With “proportion satisfactory” on the ordinate, graphs like Figure 2 were drawn using data in the Taylor-Russell tables. A set of curves for r , r^2 , and E was plotted for

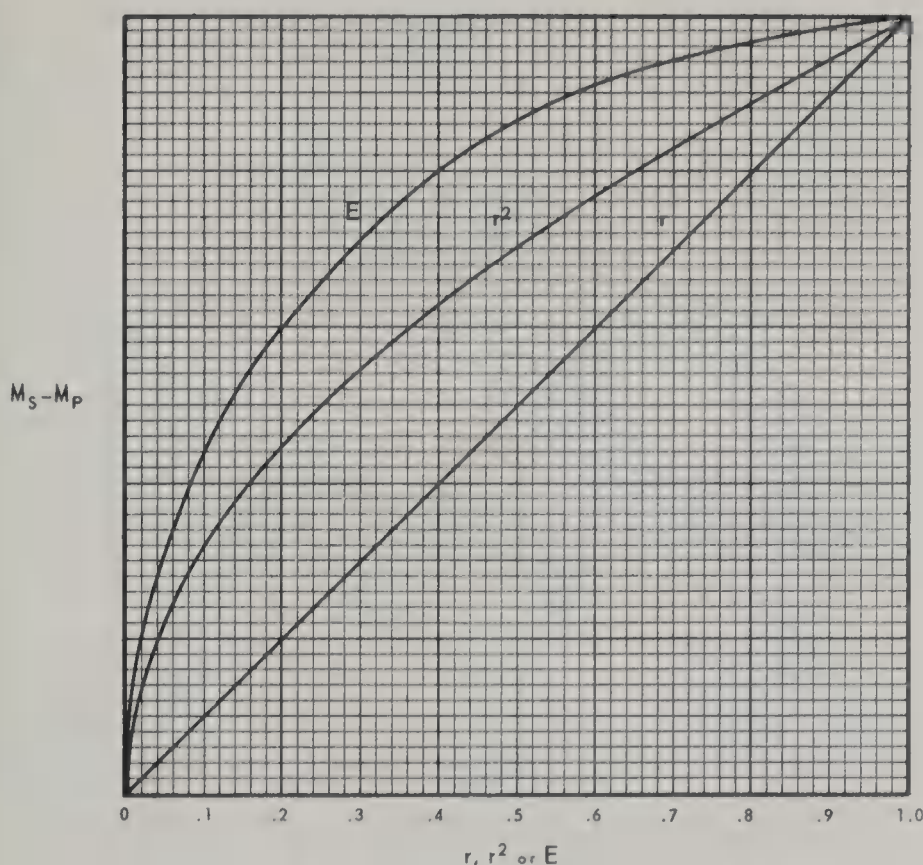


FIG. 2. Increase of the criterion mean as a function of r , r^2 , and E .

each of the 121 combinations of base rate and selection ratio in the Taylor-Russell tables. When r varies above .70, which is rare in selection testing, there is no consistent difference in terms of linearity for r , r^2 , and E . In contrast, when r varies below .70, the curve for r in most of the 121 sets of curves is decidedly more linear than the curve for r^2 and E . Usually, the set looks much like the set in Figure 2.

As a rigorous test of relative linearity, the correlation between each of the three indexes and proportion satisfactory was computed. Fifteen curves were not tested because they are nearly horizontal—that is, proportion satisfactory changes by three or less points between r values of zero and 1.0. Of the 106 curves tested, 101 cases yielded a higher correlation for r than for r^2 or E , two yielded equal correlations, and three yielded lower correlations for r . Therefore, proportion satisfactory is more nearly a linear function of r than of r^2 or E .

The curves relating r and proportion satisfactory are very nearly linear when r varies

between zero and .70. The correlation between r and the proportion satisfactory is .95 or greater for 114 of the 121 curves. The seven exceptions have extreme selection ratios and proportion satisfactory values of .90 or .95.

Thus, when proportion satisfactory is the measure of practical significance, the r scale is more meaningful for selection test evaluation than either the r^2 or the E scale.

PROPORTIONS IN FINER CRITERION CATEGORIES

The Taylor-Russell tables are not directly applicable when there are more than two important criterion categories, which is usually true when the criterion is school grades or supervisor ratings. When this is the case, the proportion of selectees expected in each category can be computed using the United States Department of Commerce (1959) tables of the bivariate normal distribution. Each combination of selection ratio and r yields a set of expected proportions.

Table 1 presents the results of the computations for 18 criterion categories. It gives

TABLE 1
PROPORTION EXPECTED IN 18 CRITERION CATEGORIES FOR GIVEN VALUES OF *r* AND SELECTION RATIO

| Criterion categories | <i>r</i> | | | | | | | | | | |
|------------------------|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | .0 | .1 | .2 | .3 | .4 | .5 | .6 | .7 | .8 | .9 | 1.0 |
| Selection ratio: .0287 | | | | | | | | | | | |
| 2.3-∞ | 011 | 017 | 031 | 045 | 066 | 094 | 129 | 174 | 233 | 307 | 373 |
| 1.9-2.3 | 018 | 028 | 038 | 059 | 077 | 101 | 132 | 167 | 216 | 296 | 627 |
| 1.4-1.9 | 052 | 073 | 098 | 125 | 157 | 188 | 223 | 261 | 289 | 296 | 000 |
| 1.0-1.4 | 078 | 101 | 122 | 143 | 164 | 181 | 188 | 185 | 160 | 084 | 000 |
| 0.7-1.0 | 083 | 098 | 115 | 122 | 129 | 129 | 119 | 098 | 059 | 014 | 000 |
| 0.5-0.7 | 066 | 077 | 080 | 084 | 080 | 073 | 063 | 045 | 024 | 004 | 000 |
| 0.3-0.5 | 074 | 077 | 080 | 077 | 073 | 063 | 049 | 028 | 011 | 000 | 000 |
| 0.1-0.3 | 078 | 080 | 077 | 073 | 066 | 052 | 038 | 021 | 004 | 000 | 000 |
| 0.0-0.1 | 040 | 042 | 038 | 035 | 028 | 021 | 011 | 007 | 004 | 000 | 000 |
| -0.1-0.0 | 040 | 042 | 035 | 031 | 028 | 017 | 011 | 007 | 004 | 000 | 000 |
| -0.3-0.1 | 078 | 070 | 066 | 052 | 042 | 028 | 017 | 004 | 000 | 000 | 000 |
| -0.5-0.3 | 074 | 066 | 056 | 045 | 031 | 021 | 011 | 004 | 000 | 000 | 000 |
| -0.7-0.5 | 066 | 059 | 045 | 035 | 021 | 014 | 004 | 004 | 000 | 000 | 000 |
| -1.0-0.7 | 083 | 066 | 049 | 035 | 021 | 011 | 004 | 000 | 000 | 000 | 000 |
| -1.4-1.0 | 078 | 056 | 042 | 024 | 014 | 004 | 004 | 000 | 000 | 000 | 000 |
| -1.9-1.4 | 052 | 035 | 021 | 011 | 007 | 004 | 000 | 000 | 000 | 000 | 000 |
| -2.3-1.9 | 018 | 011 | 004 | 004 | 000 | 000 | 000 | 000 | 000 | 000 | 000 |
| ∞-2.3 | 011 | 007 | 003 | 000 | 000 | 000 | 000 | 000 | 000 | 000 | 000 |
| Selection ratio: .0808 | | | | | | | | | | | |
| 2.3-∞ | 011 | 017 | 025 | 035 | 047 | 062 | 078 | 097 | 115 | 130 | 132 |
| 1.9-2.3 | 018 | 025 | 035 | 047 | 059 | 074 | 094 | 116 | 147 | 189 | 223 |
| 1.4-1.9 | 052 | 069 | 088 | 108 | 132 | 160 | 191 | 229 | 277 | 353 | 645 |
| 1.0-1.4 | 078 | 095 | 114 | 132 | 151 | 171 | 188 | 207 | 220 | 219 | 000 |
| 0.7-1.0 | 083 | 095 | 108 | 119 | 128 | 134 | 137 | 132 | 119 | 077 | 000 |
| 0.5-0.7 | 066 | 074 | 079 | 082 | 084 | 083 | 079 | 071 | 051 | 020 | 000 |
| 0.3-0.5 | 074 | 078 | 080 | 083 | 080 | 076 | 067 | 053 | 033 | 009 | 000 |
| 0.1-0.3 | 078 | 079 | 079 | 077 | 072 | 064 | 053 | 038 | 019 | 003 | 000 |
| 0.0-0.1 | 040 | 040 | 038 | 036 | 033 | 029 | 021 | 011 | 006 | 000 | 000 |
| -0.1-0.0 | 040 | 038 | 037 | 033 | 030 | 025 | 019 | 011 | 004 | 000 | 000 |
| -0.3-0.1 | 078 | 074 | 068 | 061 | 052 | 041 | 027 | 015 | 005 | 000 | 000 |
| -0.5-0.3 | 074 | 068 | 059 | 051 | 041 | 031 | 019 | 009 | 001 | 000 | 000 |
| -0.7-0.5 | 066 | 058 | 051 | 041 | 030 | 020 | 011 | 005 | 001 | 000 | 000 |
| -1.0-0.7 | 083 | 071 | 056 | 042 | 030 | 019 | 010 | 003 | 000 | 000 | 000 |
| -1.4-1.0 | 078 | 061 | 046 | 032 | 020 | 011 | 004 | 001 | 000 | 000 | 000 |
| -1.9-1.4 | 052 | 037 | 026 | 016 | 009 | 005 | 001 | 000 | 000 | 000 | 000 |
| -2.3-1.9 | 018 | 012 | 007 | 004 | 003 | 001 | 000 | 000 | 000 | 000 | 000 |
| ∞-2.3 | 011 | 006 | 004 | 001 | 000 | 000 | 000 | 000 | 000 | 000 | 000 |
| Selection ratio: .1587 | | | | | | | | | | | |
| 2.3-∞ | 011 | 016 | 021 | 028 | 035 | 043 | 052 | 059 | 065 | 067 | 067 |
| 1.9-2.3 | 018 | 024 | 031 | 040 | 049 | 059 | 070 | 083 | 098 | 110 | 113 |
| 1.4-1.9 | 052 | 066 | 081 | 096 | 115 | 136 | 159 | 188 | 224 | 276 | 328 |
| 1.0-1.4 | 078 | 092 | 107 | 123 | 139 | 156 | 176 | 199 | 228 | 274 | 491 |
| 0.7-1.0 | 083 | 094 | 104 | 113 | 123 | 132 | 141 | 149 | 155 | 154 | 000 |
| 0.5-0.7 | 066 | 073 | 078 | 082 | 086 | 088 | 089 | 088 | 081 | 062 | 000 |
| 0.3-0.5 | 074 | 078 | 080 | 083 | 083 | 083 | 079 | 073 | 060 | 033 | 000 |
| 0.1-0.3 | 078 | 080 | 081 | 080 | 078 | 073 | 067 | 057 | 040 | 015 | 000 |
| 0.0-0.1 | 040 | 040 | 040 | 038 | 036 | 033 | 028 | 022 | 014 | 003 | 000 |

Table 1—(Continued)

| Criterion categories | <i>r</i> | | | | | | | | | | |
|------------------------|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | .0 | .1 | .2 | .3 | .4 | .5 | .6 | .7 | .8 | .9 | 1.0 |
| Selection ratio: .1587 | | | | | | | | | | | |
| −0.1–0.0 | 040 | 039 | 038 | 036 | 033 | 030 | 025 | 019 | 011 | 002 | 000 |
| −0.3–−0.1 | 078 | 076 | 071 | 066 | 059 | 050 | 040 | 027 | 013 | 002 | 000 |
| −0.5–−0.3 | 074 | 069 | 062 | 056 | 048 | 039 | 028 | 017 | 006 | 001 | 000 |
| −0.7–−0.5 | 066 | 060 | 053 | 045 | 037 | 028 | 019 | 010 | 003 | 000 | 000 |
| −1.0–−0.7 | 083 | 073 | 062 | 050 | 038 | 027 | 016 | 007 | 001 | 000 | 000 |
| −1.4–−1.0 | 078 | 064 | 050 | 038 | 027 | 016 | 008 | 003 | 001 | 000 | 000 |
| −1.9–−1.4 | 052 | 040 | 029 | 020 | 012 | 006 | 002 | 001 | 000 | 000 | 000 |
| −2.3–−1.9 | 018 | 013 | 009 | 005 | 003 | 001 | 001 | 001 | 000 | 000 | 000 |
| −∞–−2.3 | 011 | 007 | 004 | 002 | 001 | 000 | 000 | 000 | 000 | 000 | 000 |
| Selection ratio: .2420 | | | | | | | | | | | |
| 2.3–∞ | 011 | 014 | 019 | 024 | 029 | 033 | 038 | 042 | 044 | 044 | 044 |
| 1.9–2.3 | 018 | 023 | 029 | 035 | 041 | 049 | 056 | 063 | 070 | 074 | 074 |
| 1.4–1.9 | 052 | 063 | 075 | 088 | 103 | 118 | 136 | 156 | 180 | 205 | 215 |
| 1.0–1.4 | 078 | 090 | 103 | 115 | 129 | 145 | 162 | 184 | 212 | 255 | 322 |
| 0.7–1.0 | 083 | 092 | 101 | 110 | 119 | 129 | 139 | 151 | 167 | 191 | 344 |
| 0.5–0.7 | 066 | 072 | 076 | 081 | 085 | 088 | 092 | 096 | 098 | 098 | 000 |
| 0.3–0.5 | 074 | 077 | 080 | 082 | 084 | 086 | 086 | 085 | 080 | 067 | 000 |
| 0.1–0.3 | 078 | 080 | 081 | 081 | 081 | 079 | 076 | 070 | 060 | 038 | 000 |
| 0.0–0.1 | 040 | 040 | 040 | 039 | 038 | 036 | 033 | 029 | 022 | 011 | 000 |
| −0.1–0.0 | 040 | 039 | 039 | 037 | 036 | 034 | 030 | 025 | 018 | 007 | 000 |
| −0.3–−0.1 | 078 | 076 | 072 | 069 | 064 | 057 | 049 | 039 | 025 | 007 | 000 |
| −0.5–−0.3 | 074 | 070 | 065 | 060 | 053 | 046 | 037 | 026 | 014 | 003 | 000 |
| −0.7–−0.5 | 066 | 061 | 055 | 049 | 042 | 034 | 026 | 016 | 007 | 001 | 000 |
| −1.0–−0.7 | 083 | 074 | 065 | 055 | 044 | 034 | 023 | 012 | 004 | 000 | 000 |
| −1.4–−1.0 | 078 | 066 | 055 | 043 | 032 | 022 | 012 | 005 | 001 | 000 | 000 |
| −1.9–−1.4 | 052 | 042 | 031 | 023 | 015 | 009 | 004 | 001 | 000 | 000 | 000 |
| −2.3–−1.9 | 018 | 013 | 010 | 006 | 004 | 002 | 001 | 000 | 000 | 000 | 000 |
| −∞–−1.3 | 011 | 007 | 005 | 002 | 001 | 000 | 000 | 000 | 000 | 000 | 000 |
| Selection ratio: .3085 | | | | | | | | | | | |
| 2.3–∞ | 011 | 014 | 018 | 021 | 025 | 029 | 031 | 034 | 035 | 035 | 035 |
| 1.9–2.3 | 018 | 022 | 027 | 032 | 037 | 043 | 048 | 053 | 057 | 058 | 058 |
| 1.4–1.9 | 052 | 062 | 072 | 083 | 095 | 108 | 122 | 137 | 152 | 166 | 169 |
| 1.0–1.4 | 078 | 089 | 100 | 111 | 124 | 137 | 152 | 171 | 195 | 227 | 253 |
| 0.7–1.0 | 083 | 091 | 099 | 108 | 116 | 125 | 136 | 149 | 166 | 195 | 270 |
| 0.5–0.7 | 066 | 071 | 075 | 079 | 084 | 088 | 093 | 098 | 105 | 115 | 216 |
| 0.3–0.5 | 074 | 077 | 080 | 082 | 085 | 087 | 089 | 090 | 091 | 089 | 000 |
| 0.1–0.3 | 078 | 080 | 081 | 081 | 082 | 082 | 080 | 078 | 073 | 059 | 000 |
| 0.0–0.1 | 040 | 040 | 040 | 040 | 039 | 038 | 036 | 033 | 029 | 019 | 000 |
| −0.1–0.0 | 040 | 040 | 039 | 038 | 037 | 035 | 028 | 029 | 024 | 014 | 000 |
| −0.3–−0.1 | 078 | 076 | 074 | 071 | 067 | 062 | 055 | 054 | 034 | 015 | 000 |
| −0.5–−0.3 | 074 | 070 | 066 | 061 | 056 | 050 | 043 | 026 | 021 | 007 | 000 |
| −0.7–−0.5 | 066 | 062 | 057 | 052 | 045 | 039 | 031 | 022 | 011 | 002 | 000 |
| −1.0–−0.7 | 083 | 075 | 067 | 058 | 049 | 039 | 028 | 017 | 007 | 001 | 000 |
| −1.4–−1.0 | 078 | 067 | 057 | 047 | 037 | 026 | 017 | 008 | 002 | 000 | 000 |
| −1.9–−1.4 | 052 | 043 | 034 | 026 | 018 | 011 | 006 | 002 | 000 | 000 | 000 |
| −2.3–−1.9 | 018 | 014 | 010 | 007 | 004 | 002 | 001 | 000 | 000 | 000 | 000 |
| −∞–−2.3 | 011 | 008 | 005 | 003 | 002 | 001 | 000 | 000 | 000 | 000 | 000 |

Table 1—(Continued)

| Criterion categories | <i>r</i> | | | | | | | | | | |
|------------------------|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | .0 | .1 | .2 | .3 | .4 | .5 | .6 | .7 | .8 | .9 | 1.0 |
| Selection ratio: .3821 | | | | | | | | | | | |
| 2.3-∞ | 011 | 014 | 016 | 020 | 022 | 025 | 026 | 027 | 028 | 028 | 028 |
| 1.9-2.3 | 018 | 022 | 026 | 030 | 034 | 038 | 041 | 045 | 046 | 047 | 047 |
| 1.4-1.9 | 052 | 061 | 069 | 079 | 088 | 098 | 109 | 119 | 129 | 136 | 136 |
| 1.0-1.4 | 078 | 087 | 097 | 107 | 117 | 129 | 142 | 157 | 175 | 195 | 204 |
| 0.7-1.0 | 083 | 090 | 098 | 105 | 113 | 121 | 131 | 143 | 160 | 185 | 218 |
| 0.5-0.7 | 066 | 070 | 074 | 079 | 082 | 087 | 092 | 099 | 107 | 123 | 174 |
| 0.3-0.5 | 074 | 076 | 079 | 082 | 085 | 087 | 091 | 094 | 099 | 107 | 193 |
| 0.1-0.3 | 078 | 080 | 081 | 082 | 083 | 084 | 084 | 084 | 084 | 080 | 000 |
| 0.0-0.1 | 040 | 040 | 040 | 040 | 040 | 039 | 038 | 037 | 035 | 029 | 000 |
| -0.1-0.0 | 040 | 040 | 039 | 039 | 038 | 037 | 036 | 034 | 030 | 022 | 000 |
| -0.3--0.1 | 078 | 076 | 075 | 072 | 069 | 066 | 061 | 055 | 046 | 028 | 000 |
| -0.5--0.3 | 074 | 071 | 067 | 064 | 059 | 054 | 048 | 040 | 029 | 013 | 000 |
| -0.7--0.5 | 066 | 062 | 058 | 053 | 049 | 043 | 036 | 028 | 017 | 005 | 000 |
| -1.0--0.7 | 083 | 076 | 069 | 062 | 053 | 044 | 034 | 023 | 012 | 002 | 000 |
| -1.4--1.0 | 078 | 069 | 059 | 050 | 041 | 031 | 021 | 012 | 004 | 000 | 000 |
| -1.9--1.4 | 052 | 044 | 036 | 028 | 021 | 014 | 008 | 003 | 001 | 000 | 000 |
| -2.3--1.9 | 018 | 014 | 011 | 008 | 005 | 003 | 001 | 001 | 000 | 000 | 000 |
| -∞--2.3 | 011 | 008 | 006 | 004 | 002 | 001 | 000 | 000 | 000 | 000 | 000 |
| Selection ratio: .4602 | | | | | | | | | | | |
| 2.3-∞ | 011 | 013 | 016 | 018 | 020 | 021 | 023 | 023 | 023 | 023 | 023 |
| 1.9-2.3 | 018 | 021 | 024 | 028 | 031 | 034 | 036 | 038 | 039 | 039 | 039 |
| 1.4-1.9 | 052 | 059 | 067 | 074 | 082 | 090 | 097 | 105 | 110 | 113 | 113 |
| 1.0-1.4 | 078 | 086 | 094 | 103 | 112 | 121 | 132 | 143 | 156 | 167 | 169 |
| 0.7-1.0 | 083 | 089 | 096 | 102 | 109 | 117 | 126 | 136 | 150 | 169 | 181 |
| 0.5-0.7 | 066 | 070 | 073 | 077 | 081 | 086 | 091 | 097 | 107 | 122 | 145 |
| 0.3-0.5 | 074 | 076 | 079 | 081 | 084 | 087 | 091 | 096 | 103 | 115 | 160 |
| 0.1-0.3 | 078 | 077 | 081 | 082 | 083 | 085 | 087 | 088 | 092 | 097 | 170 |
| 0.0-0.1 | 040 | 040 | 040 | 040 | 040 | 040 | 040 | 040 | 040 | 039 | 000 |
| -0.1-0.0 | 040 | 040 | 039 | 039 | 039 | 038 | 038 | 037 | 035 | 032 | 000 |
| -0.3--0.1 | 078 | 077 | 075 | 073 | 072 | 069 | 067 | 063 | 057 | 045 | 000 |
| -0.5--0.3 | 074 | 071 | 068 | 065 | 062 | 059 | 054 | 048 | 039 | 024 | 000 |
| -0.7--0.5 | 066 | 063 | 060 | 056 | 051 | 047 | 041 | 034 | 024 | 011 | 000 |
| -1.0--0.7 | 083 | 077 | 071 | 064 | 057 | 050 | 041 | 030 | 018 | 005 | 000 |
| -1.4--1.0 | 078 | 070 | 062 | 054 | 045 | 036 | 027 | 017 | 007 | 001 | 000 |
| -1.9--1.4 | 052 | 045 | 038 | 031 | 024 | 017 | 010 | 005 | 001 | 000 | 000 |
| -2.3--1.9 | 018 | 015 | 012 | 009 | 006 | 004 | 002 | 001 | 000 | 000 | 000 |
| -∞--2.3 | 011 | 008 | 006 | 004 | 002 | 001 | 000 | 000 | 000 | 000 | 000 |
| Selection ratio: .5398 | | | | | | | | | | | |
| 2.3-∞ | 011 | 013 | 015 | 016 | 018 | 019 | 019 | 020 | 020 | 020 | 020 |
| 1.9-2.3 | 018 | 021 | 023 | 026 | 028 | 030 | 032 | 033 | 033 | 033 | 033 |
| 1.4-1.9 | 052 | 058 | 065 | 070 | 076 | 082 | 088 | 092 | 095 | 097 | 097 |
| 1.0-1.4 | 078 | 085 | 092 | 099 | 106 | 114 | 122 | 130 | 138 | 144 | 144 |
| 0.7-1.0 | 083 | 089 | 094 | 100 | 105 | 112 | 120 | 128 | 139 | 150 | 154 |
| 0.5-0.7 | 066 | 070 | 072 | 076 | 080 | 083 | 088 | 094 | 102 | 114 | 123 |
| 0.3-0.5 | 074 | 076 | 078 | 082 | 083 | 087 | 091 | 096 | 103 | 116 | 136 |
| 0.1-0.3 | 078 | 079 | 081 | 081 | 084 | 086 | 088 | 091 | 097 | 107 | 145 |
| 0.0-0.1 | 040 | 040 | 040 | 040 | 041 | 041 | 042 | 042 | 044 | 047 | 074 |
| -0.1-0.0 | 040 | 040 | 040 | 040 | 040 | 039 | 040 | 040 | 040 | 041 | 074 |

Table 1—(Continued)

| Criterion categories | <i>r</i> | | | | | | | | | | |
|------------------------|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | .0 | .1 | .2 | .3 | .4 | .5 | .6 | .7 | .8 | .9 | 1.0 |
| Selection ratio: .5398 | | | | | | | | | | | |
| −0.3−0.1 | 078 | 077 | 076 | 075 | 074 | 072 | 071 | 069 | 067 | 062 | 000 |
| −0.5−0.3 | 074 | 072 | 069 | 067 | 065 | 062 | 059 | 055 | 049 | 038 | 000 |
| −0.7−0.5 | 066 | 064 | 061 | 057 | 054 | 050 | 046 | 040 | 032 | 020 | 000 |
| −1.0−0.7 | 083 | 078 | 073 | 067 | 061 | 055 | 047 | 038 | 026 | 011 | 000 |
| −1.4−1.0 | 078 | 071 | 064 | 057 | 049 | 041 | 032 | 022 | 012 | 002 | 000 |
| −1.9−1.4 | 052 | 046 | 040 | 033 | 027 | 020 | 014 | 007 | 003 | 000 | 000 |
| −2.3−1.9 | 018 | 015 | 013 | 010 | 007 | 005 | 003 | 001 | 000 | 000 | 000 |
| −∞−2.3 | 011 | 009 | 006 | 005 | 003 | 003 | 001 | 000 | 000 | 000 | 000 |
| Selection ratio: .6179 | | | | | | | | | | | |
| 2.3−∞ | 011 | 012 | 014 | 015 | 016 | 017 | 017 | 017 | 017 | 017 | 017 |
| 1.9−2.3 | 018 | 020 | 022 | 024 | 026 | 027 | 028 | 029 | 029 | 029 | 029 |
| 1.4−1.9 | 052 | 057 | 062 | 067 | 072 | 076 | 080 | 082 | 084 | 084 | 084 |
| 1.0−1.4 | 078 | 084 | 089 | 095 | 101 | 107 | 113 | 119 | 124 | 126 | 126 |
| 0.7−1.0 | 083 | 088 | 092 | 097 | 102 | 108 | 114 | 120 | 128 | 134 | 135 |
| 0.5−0.7 | 066 | 069 | 072 | 075 | 078 | 081 | 086 | 091 | 097 | 105 | 108 |
| 0.3−0.5 | 074 | 075 | 078 | 080 | 083 | 086 | 089 | 095 | 101 | 111 | 119 |
| 0.1−0.3 | 078 | 079 | 080 | 082 | 084 | 086 | 089 | 092 | 098 | 109 | 126 |
| 0.0−0.1 | 040 | 040 | 040 | 041 | 041 | 042 | 042 | 044 | 046 | 051 | 064 |
| −0.1−0.0 | 040 | 040 | 040 | 040 | 040 | 040 | 041 | 042 | 043 | 046 | 064 |
| −0.3−0.1 | 078 | 077 | 076 | 076 | 075 | 075 | 074 | 074 | 075 | 077 | 126 |
| −0.5−0.3 | 074 | 072 | 070 | 069 | 067 | 065 | 063 | 061 | 058 | 053 | 000 |
| −0.7−0.5 | 066 | 064 | 062 | 059 | 057 | 054 | 051 | 047 | 041 | 032 | 000 |
| −1.0−0.7 | 083 | 079 | 074 | 070 | 065 | 060 | 054 | 046 | 036 | 020 | 000 |
| −1.4−1.0 | 078 | 072 | 066 | 060 | 054 | 046 | 038 | 029 | 018 | 006 | 000 |
| −1.9−1.4 | 052 | 047 | 041 | 036 | 030 | 024 | 017 | 011 | 004 | 001 | 000 |
| −2.3−1.9 | 018 | 016 | 013 | 011 | 008 | 006 | 004 | 002 | 001 | 000 | 000 |
| −∞−2.3 | 011 | 009 | 007 | 005 | 004 | 002 | 001 | 000 | 000 | 000 | 000 |
| Selection ratio: .6915 | | | | | | | | | | | |
| 2.3−∞ | 011 | 012 | 013 | 014 | 015 | 015 | 015 | 015 | 015 | 015 | 015 |
| 1.9−2.3 | 018 | 020 | 022 | 023 | 024 | 025 | 026 | 026 | 026 | 026 | 026 |
| 1.4−1.9 | 052 | 056 | 060 | 064 | 067 | 070 | 073 | 075 | 075 | 075 | 075 |
| 1.0−1.4 | 078 | 083 | 087 | 092 | 096 | 101 | 105 | 109 | 112 | 113 | 113 |
| 0.7−1.0 | 083 | 087 | 091 | 095 | 099 | 103 | 108 | 113 | 117 | 120 | 121 |
| 0.5−0.7 | 066 | 069 | 071 | 073 | 076 | 079 | 082 | 087 | 091 | 095 | 096 |
| 0.3−0.5 | 074 | 075 | 077 | 079 | 081 | 084 | 088 | 092 | 097 | 104 | 106 |
| 0.1−0.3 | 078 | 079 | 080 | 081 | 083 | 085 | 088 | 092 | 098 | 106 | 113 |
| 0.0−0.1 | 040 | 040 | 040 | 041 | 041 | 042 | 043 | 045 | 047 | 052 | 058 |
| −0.1−0.0 | 040 | 040 | 040 | 040 | 040 | 041 | 042 | 043 | 045 | 049 | 058 |
| −0.3−0.1 | 078 | 077 | 077 | 077 | 076 | 077 | 077 | 078 | 081 | 087 | 113 |
| −0.5−0.3 | 074 | 072 | 071 | 070 | 069 | 068 | 067 | 066 | 066 | 067 | 106 |
| −0.7−0.5 | 066 | 064 | 063 | 061 | 059 | 057 | 055 | 052 | 049 | 045 | 000 |
| −1.0−0.7 | 083 | 080 | 076 | 073 | 069 | 065 | 060 | 054 | 047 | 034 | 000 |
| −1.4−1.0 | 078 | 073 | 068 | 063 | 058 | 052 | 045 | 036 | 026 | 012 | 000 |
| −1.9−1.4 | 052 | 048 | 043 | 038 | 033 | 027 | 021 | 015 | 008 | 001 | 000 |
| −2.3−1.9 | 018 | 016 | 014 | 012 | 009 | 007 | 005 | 003 | 001 | 000 | 000 |
| −∞−2.3 | 011 | 009 | 008 | 006 | 004 | 003 | 001 | 000 | 000 | 000 | 000 |

Table 1—(Continued)

| Criterion categories | <i>r</i> | | | | | | | | | | |
|------------------------|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | .0 | .1 | .2 | .3 | .4 | .5 | .6 | .7 | .8 | .9 | 1.0 |
| Selection ratio: .7580 | | | | | | | | | | | |
| 2.3-∞ | 011 | 012 | 013 | 013 | 014 | 014 | 014 | 014 | 014 | 014 | 014 |
| 1.9-2.3 | 018 | 020 | 021 | 022 | 023 | 023 | 024 | 024 | 024 | 024 | 024 |
| 1.4-1.9 | 052 | 055 | 059 | 061 | 064 | 066 | 067 | 068 | 069 | 069 | 069 |
| 1.0-1.4 | 078 | 082 | 085 | 089 | 093 | 096 | 099 | 101 | 102 | 103 | 103 |
| 0.7-1.0 | 083 | 086 | 089 | 093 | 096 | 099 | 103 | 106 | 109 | 110 | 110 |
| 0.5-0.7 | 066 | 068 | 070 | 072 | 074 | 077 | 080 | 083 | 086 | 088 | 088 |
| 0.3-0.5 | 074 | 075 | 076 | 078 | 080 | 083 | 085 | 089 | 093 | 096 | 097 |
| 0.1-0.3 | 078 | 079 | 080 | 081 | 083 | 085 | 087 | 091 | 095 | 101 | 103 |
| 0.0-0.1 | 040 | 040 | 040 | 041 | 041 | 042 | 043 | 045 | 047 | 050 | 053 |
| -0.1-0.0 | 040 | 040 | 040 | 040 | 041 | 041 | 042 | 043 | 045 | 049 | 053 |
| -0.3--0.1 | 078 | 078 | 077 | 077 | 077 | 078 | 079 | 081 | 084 | 091 | 103 |
| -0.5--0.3 | 074 | 072 | 072 | 071 | 070 | 070 | 070 | 070 | 072 | 076 | 097 |
| -0.7--0.5 | 066 | 065 | 064 | 062 | 061 | 060 | 058 | 057 | 057 | 057 | 088 |
| -1.0--0.7 | 083 | 081 | 078 | 075 | 072 | 069 | 065 | 062 | 057 | 049 | 000 |
| -1.4--1.0 | 078 | 074 | 070 | 066 | 062 | 057 | 051 | 044 | 035 | 021 | 000 |
| -1.9--1.4 | 052 | 049 | 045 | 041 | 036 | 031 | 025 | 019 | 011 | 003 | 000 |
| -2.3--1.9 | 018 | 016 | 015 | 014 | 011 | 008 | 006 | 004 | 002 | 000 | 000 |
| -∞--2.3 | 011 | 010 | 008 | 006 | 005 | 003 | 002 | 001 | 000 | 000 | 000 |
| Selection ratio: .8413 | | | | | | | | | | | |
| 2.3-∞ | 011 | 011 | 012 | 012 | 013 | 013 | 013 | 013 | 013 | 013 | 013 |
| 1.9-2.3 | 018 | 019 | 020 | 020 | 021 | 021 | 021 | 021 | 021 | 021 | 021 |
| 1.4-1.9 | 052 | 054 | 057 | 058 | 060 | 061 | 062 | 062 | 062 | 062 | 062 |
| 1.0-1.4 | 078 | 081 | 083 | 086 | 088 | 090 | 091 | 092 | 093 | 093 | 093 |
| 0.7-1.0 | 083 | 085 | 087 | 090 | 092 | 094 | 096 | 098 | 099 | 099 | 099 |
| 0.5-0.7 | 066 | 068 | 069 | 071 | 072 | 074 | 076 | 077 | 079 | 079 | 079 |
| 0.3-0.5 | 074 | 075 | 076 | 077 | 079 | 080 | 082 | 084 | 086 | 087 | 088 |
| 0.1-0.3 | 078 | 079 | 079 | 080 | 082 | 083 | 085 | 088 | 090 | 093 | 093 |
| 0.0-0.1 | 040 | 040 | 040 | 041 | 041 | 042 | 043 | 044 | 045 | 047 | 047 |
| -0.1-0.0 | 040 | 040 | 040 | 040 | 041 | 041 | 042 | 043 | 045 | 047 | 047 |
| -0.3--0.1 | 078 | 078 | 078 | 078 | 078 | 079 | 080 | 082 | 085 | 090 | 093 |
| -0.5--0.3 | 074 | 073 | 072 | 072 | 072 | 072 | 073 | 074 | 076 | 081 | 088 |
| -0.7--0.5 | 066 | 065 | 064 | 064 | 063 | 062 | 062 | 063 | 064 | 067 | 079 |
| -1.0--0.7 | 083 | 081 | 079 | 078 | 076 | 074 | 072 | 071 | 070 | 070 | 099 |
| -1.4--1.0 | 078 | 075 | 072 | 069 | 066 | 063 | 059 | 055 | 050 | 041 | 000 |
| -1.9--1.4 | 052 | 050 | 047 | 044 | 040 | 036 | 032 | 026 | 020 | 010 | 000 |
| -2.3--1.9 | 018 | 017 | 016 | 014 | 012 | 010 | 008 | 006 | 003 | 001 | 000 |
| -∞--2.3 | 011 | 010 | 009 | 007 | 006 | 004 | 003 | 002 | 000 | 000 | 000 |
| Selection ratio: .9192 | | | | | | | | | | | |
| 2.3-∞ | 011 | 011 | 011 | 012 | 012 | 012 | 012 | 012 | 012 | 012 | 012 |
| 1.9-2.3 | 018 | 019 | 019 | 019 | 019 | 020 | 020 | 020 | 020 | 020 | 020 |
| 1.4-1.9 | 052 | 053 | 054 | 055 | 056 | 057 | 057 | 057 | 057 | 057 | 057 |
| 1.0-1.4 | 078 | 079 | 081 | 082 | 083 | 084 | 084 | 085 | 085 | 085 | 085 |
| 0.7-1.0 | 083 | 084 | 086 | 087 | 088 | 089 | 090 | 090 | 091 | 091 | 091 |
| 0.5-0.7 | 066 | 067 | 068 | 069 | 070 | 071 | 071 | 072 | 072 | 072 | 072 |
| 0.3-0.5 | 074 | 074 | 075 | 076 | 077 | 078 | 078 | 079 | 080 | 080 | 080 |
| 0.1-0.3 | 078 | 078 | 079 | 080 | 080 | 081 | 083 | 084 | 085 | 085 | 085 |
| 0.0-0.1 | 040 | 040 | 040 | 040 | 041 | 041 | 042 | 042 | 043 | 043 | 043 |
| -0.1-0.0 | 040 | 040 | 040 | 040 | 040 | 041 | 041 | 042 | 043 | 043 | 043 |

Table 1—(Continued)

| Criterion categories | <i>r</i> | | | | | | | | | | |
|------------------------|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | .0 | .1 | .2 | .3 | .4 | .5 | .6 | .7 | .8 | .9 | 1.0 |
| Selection ratio: .9192 | | | | | | | | | | | |
| −0.3—0.1 | 078 | 078 | 078 | 078 | 079 | 079 | 080 | 082 | 083 | 085 | 085 |
| −0.5—0.3 | 074 | 073 | 073 | 073 | 073 | 073 | 074 | 075 | 077 | 079 | 080 |
| −0.7—0.5 | 066 | 066 | 065 | 065 | 065 | 065 | 065 | 066 | 068 | 071 | 072 |
| −1.0—0.7 | 083 | 082 | 081 | 080 | 079 | 079 | 079 | 079 | 080 | 084 | 091 |
| −1.4—1.0 | 078 | 076 | 075 | 073 | 072 | 070 | 068 | 067 | 065 | 066 | 085 |
| −1.9—1.4 | 052 | 051 | 049 | 047 | 045 | 043 | 040 | 037 | 032 | 026 | 000 |
| −2.3—1.9 | 018 | 017 | 017 | 015 | 014 | 013 | 011 | 009 | 007 | 003 | 000 |
| −∞—2.3 | 011 | 010 | 009 | 009 | 008 | 006 | 005 | 003 | 002 | 000 | 000 |
| Selection ratio: .9713 | | | | | | | | | | | |
| 2.3—∞ | 011 | 011 | 011 | 011 | 011 | 011 | 011 | 011 | 011 | 011 | 011 |
| 1.9—2.3 | 018 | 018 | 018 | 018 | 019 | 019 | 019 | 019 | 019 | 019 | 019 |
| 1.4—1.9 | 052 | 053 | 053 | 053 | 053 | 054 | 054 | 054 | 054 | 054 | 054 |
| 1.0—1.4 | 078 | 079 | 079 | 080 | 080 | 080 | 080 | 080 | 080 | 080 | 080 |
| 0.7—1.0 | 083 | 084 | 084 | 085 | 085 | 086 | 086 | 086 | 086 | 086 | 086 |
| 0.5—0.7 | 066 | 067 | 067 | 067 | 068 | 068 | 068 | 068 | 069 | 069 | 069 |
| 0.3—0.5 | 074 | 074 | 074 | 074 | 075 | 075 | 076 | 076 | 076 | 076 | 076 |
| 0.1—0.3 | 078 | 078 | 079 | 079 | 079 | 080 | 080 | 080 | 080 | 080 | 080 |
| 0.0—0.1 | 040 | 040 | 040 | 040 | 040 | 041 | 041 | 041 | 041 | 041 | 041 |
| −0.1—0.0 | 040 | 040 | 040 | 040 | 040 | 040 | 041 | 041 | 041 | 041 | 041 |
| −0.3—0.1 | 078 | 079 | 078 | 078 | 079 | 079 | 079 | 080 | 080 | 080 | 080 |
| −0.5—0.3 | 074 | 073 | 074 | 074 | 074 | 074 | 074 | 075 | 076 | 076 | 076 |
| −0.7—0.5 | 066 | 066 | 066 | 066 | 066 | 066 | 067 | 067 | 068 | 068 | 069 |
| −1.0—0.7 | 083 | 083 | 082 | 082 | 082 | 082 | 082 | 083 | 084 | 085 | 086 |
| −1.4—1.0 | 078 | 077 | 077 | 076 | 075 | 075 | 075 | 075 | 076 | 078 | 080 |
| −1.9—1.4 | 052 | 052 | 051 | 050 | 049 | 048 | 047 | 046 | 045 | 045 | 054 |
| −2.3—1.9 | 018 | 018 | 017 | 017 | 016 | 016 | 015 | 014 | 012 | 010 | 000 |
| −∞—2.3 | 011 | 011 | 010 | 010 | 009 | 008 | 007 | 006 | 004 | 002 | 000 |

the proportions expected in each criterion category when *r* is .0, .1, .2, .3, .4, .5, .6, .7, .8, .9, or 1.0. The criterion categories are listed in the first column. Each sigma value that separates two categories is exact to the tenths digit, and therefore required no interpolation in the bivariate tables. The sigma values divide the normal curve into areas whose proportions are given in the second column where *r* is zero. With the exception of 2.3 and 0.0, these sigma values were also used on the selection test scale to define the cutoffs and, therefore, selection ratios. The Table consists of 14 sections, one for each selection ratio.³

³ Although the selection of sigma values in the bivariate tables was somewhat arbitrary, an effort

Table 1 shows the practical significance of specific changes in *r* and the selection ratio. In general, as *r* increases, the proportions in higher criterion categories increase, while the proportions in low criterion categories decrease. To compare *r*, *r*², and *E*, some of the rows were combined, yielding 10 standard deviation categories with the following boundaries: −∞, −1.9, −1.4, −1.0, −0.5, 0.0, 0.5, 1.0, 1.4, 1.9, and ∞. With 14 selection ratios under consideration, there are now 140 relationships between *r* and proportion expected.

was made to make Table 1 most useful by choosing sigma values that (a) divide the normal curve into a large number of categories with practically significant proportions, and (b) represent selection ratios that round to even percentages with a maximum error of 2/10 of 1%.

The linearity of each relationship below an r of .70 was estimated by computing the correlation between r and proportion expected. For 129 of the 140 relationships the correlation between r and proportion expected is equal to or greater than .95, which indicates a high degree of linearity. When r was transformed to r^2 , the correlation decreased in 110 of the cases, was unchanged in 8 cases, and increased in 22 cases. The correlation for the E transformation of r decreased in 111 of the cases. Thus, proportion expected is more nearly a linear function of r than of r^2 or E . Moreover, 16 of the 22 cases in which r^2 is more linear than r pertain to the two categories separated by the mean—the two least likely to be of practical significance in applied situations.

Since proportion expected in criterion categories is more nearly a linear function of r than it is of r^2 or E , r is the more useful index for this measure of practical significance.

DISCUSSION

Although Table 1 gives the proportion of selectees expected in 18 categories, expected proportions in larger categories can be determined by adding vertically adjacent numbers in the table. It is even possible to determine the proportion for a dichotomy by adding the numbers above or below a category boundary in Table 1 and subtracting the sum from 1.00. Eighteen dichotomies are possible, whereas only 11 are possible using the Taylor-Russell tables.

Another use of the table is to forecast the practical significance of specific changes in r and/or the selection ratio. By referring to the table, one can see in each criterion category the expected consequences of improving the test, that is, raising r a certain amount, or the expected consequences of changing the selection ratio a certain amount. Similarly, if a shorter test is desired even though a lower r can be expected, one can see the practical consequences of a specific reduction in r .

Table 1 can also be used to estimate the probability that observed data are from a bivariate normal population, an assumption that is often necessary when interpreting sam-

ple statistics. Obtained proportions in the criterion categories can be compared to the expected proportions in Table 1 using the χ^2 statistic (Guilford, 1965, p. 246). If the computed χ^2 value is statistically significant, this is evidence that either the sample is biased or the population is not bivariate normal. This is especially important when correcting r for restriction of the range since most procedures for correcting r assume bivariate normality in the population and a random sample from the correlation surface above the selection test cutoff.

One limitation must be placed on the conclusions in this paper regarding the relative utility of r , r^2 , and E as indexes of practical significance: The means and proportions that indicate practical significance pertain only to selectees, that is, applicants above the selection test cutoff. The criterion distribution, whether actual or hypothetical, below this cutoff is assumed to have no practical significance for the institution using the selection test. When the criterion is dichotomous, this assumption is equivalent to assigning a zero utility to two of the four decision-outcome combinations in the fourfold table, namely to reject-unsatisfactory and reject-satisfactory ("false positives"). This is consistent with the Cronbach and Gleser (1965) treatment of selection testing.

REFERENCES

- BROGDEN, H. E. On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology*, 1946, **37**, 65-76.
- CRONBACH, L. J., & GLESER, G. C. *Psychological tests and personnel decisions*. (2nd ed.) Urbana: University of Illinois Press, 1965.
- DUNNETTE, M. D. *Personnel selection and placement*. Belmont, California: Wadsworth, 1966.
- GHISELLI, E. E., & BROWN, C. W. *Personnel and industrial psychology*. New York: McGraw-Hill, 1948.
- GUILFORD, J. P. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.
- GUILFORD, J. P. *Fundamental statistics in psychology and education*. (4th ed.) New York: McGraw-Hill, 1965.
- GUION, R. M. *Personnel testing*. New York: McGraw-Hill, 1965.
- HORST, P. *Psychological measurement and prediction*. Belmont, California: Wadsworth, 1966.
- NUNNALLY, J. C., JR. *Tests and measurements*. New York: McGraw-Hill, 1959.

TAYLOR, H. C., & RUSSELL, J. T. The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology*, 1939, 23, 565-578.

THORNDIKE, R. L. *Personnel selection*. New York: Wiley, 1949.

UNITED STATES DEPARTMENT OF COMMERCE. *Tables of the bivariate normal distribution function and related functions*. Washington, D. C.: United States Government Printing Office, 1959.

(Received July 26, 1968)

GROUP AND INDIVIDUAL EFFECTS IN PROBLEM SOLVING¹

GEORGE S. ROTTER²

AND

STEPHEN M. PORTUGAL

Montclair State College

Long Island University

It was hypothesized that when working on a single problem, the combination of individual and group sessions would lead to more solutions than only individual or only group sessions. Dividing 128 male and female Ss into 32 real and nominal work groups, the hypothesis was not confirmed. Instead, the individual production of ideas was found to be superior to either group production or the combination of group and individual production ($p < .05$). In general, the production of ideas appears to be simply related to the proportion of time spent working alone. Possible explanations are discussed and areas for future research are presented.

In the ongoing debate over the relative merits of group and individual problem solving, Taylor, Berry, and Block (1958) made a fundamental methodological contribution when they divided into nominal four-person groups the separate performances of 48 individuals. After eliminating duplicate answers to problems administered under brainstorming conditions, Osborn's (1953, pp. 297-307), findings indicated that the nominal groups, that is, individual performance, were clearly superior to an equal number of real interacting groups in the production of ideas. Dunnette, Campbell, and Jaastad (1963) and Meadow, Parnes, and Reese (1959) had the same Ss engage in both group and individual sessions and their results replicated essentially those of Taylor et al.

The explanations for the finding usually are phrased in terms of an inhibitory effect created by being surrounded by other people or that the public expression of an idea in a group causes all the members to think along the same lines, leading to the duplication of ideas. It is posited, however, that individual and group conditions in brainstorming or problem solving in general have their own unique contributions to offer. Dunnette, Campbell, and Jaastad (1963) did have the same Ss work under both group and individual circumstances. Their findings are limited, however, because the shift in working conditions

was accompanied by a shift in problems. The present study allows its Ss to work on the same problem under both individual and group conditions.

Specifically, it is hypothesized that the number of solutions to a problem should be greater where S works in both a group and an individual setting than where he works in either a group or an individual setting. Because of the assumed and generally reported superiority of individual over group conditions in problem solving, it is hypothesized further that for all conditions the number of solutions to a problem should be greater when working individually than when working with a group.

METHOD

Subjects

Recruited from introductory psychology courses at Long Island University were 64 males and 64 females. Aside from their sex and the restrictions imposed by the hours they were available, assignment to the various conditions was random.

Materials

Two problems, each having numerous answers, were utilized.

Tourist problem. Each year a great many American tourists go to visit Europe. But now suppose that our country wished to get many more European tourists to come to visit America during their vacations. What steps can you suggest to get more European tourists to come to this country?

Education problem. Because of the rapidly increasing birthrate beginning in the 1940s, it is now clear that by 1980 public school enrollment will be very much greater than it is today. In fact, it has been estimated that if the student-teacher ratio were to be maintained at what it is today, 50% of all in-

¹ This paper is based, in part, on a Master's thesis by the author submitted to Long Island University.

² Requests for reprints should be sent to George S. Rotter, Psychology Department, Montclair State College, Upper Montclair, New Jersey 07043.

dividuals graduating from college would have to be induced to enter teaching. What different steps might be taken to insure that schools will continue to provide instruction at least equal in effectiveness to that now provided?

Both of these problems were originally used by Taylor et al. (1958), the only change being the revision of the dates in the second problem so that they were appropriate to the present time.

Experimental Design and Procedure

Four experimental conditions were utilized, each composed of eight same-sex groups of four members: Condition I, working individually; Condition G, working with a group; Condition I-G, working indiivdually for the first half of the session and then working with a group; Condition G-I, working with a group for the first half and then working individually. Within each condition four groups consisted of males only and four consisted of females only. Within these subdivisions, two groups worked on the Tourist problem only while the other two groups worked on the Education problem only.

Individual Conditions. Ss were told this was an experiment on creative thinking. They were read one of the problems twice and told to write down all the ideas, that is, solutions, they could think of. Quantity, not quality, was stressed. They were then sent to separate cubicles where they worked on their answers privately for 16 min.

Group Conditions. As above, the Ss were told this was an experiment on creative thinking. They were read one of the problems twice and told to write down all the ideas they could think of and heard. Again, quantity, not quality, was stressed and they were urged to avoid any manifestation of criticism or ridicule of ideas. All answers were listed by each S on his own sheet of paper. In this instance, the four-person group sat around a large table while discussing their problem for 16 min.

Individual then Group. The Ss were initially given instructions identical to those given Ss in the I condition. The exception was that they were permitted only 8 min. to work privately on their assigned problem. After this they were reassembled and given instructions identical to those administered to Ss in Condition G. They were also asked not to raise before the group the ideas produced privately. They then proceded to work as an interacting group for an additional 8 min. They used the same problem previously assigned to them as individuals.

Group then Individual. The Ss were initially given instructions identical to those given Ss in Condition G with the exception they were permitted only 8 min. to work on the problem. Following this they were told they would now work on the same problem separately and were given instructions identical to those provided in Condition I. They were asked not to use any of the same ideas raised during the group session and were sent to their respective cubicles for an additional 8 min.

All questions were answered by reiterating relevant portions of the instructions or by politely dismissing

TABLE 1
FACTORIAL ANALYSIS OF VARIANCE
OF NUMBER OF IDEAS

| Source | df | MS | F |
|----------------|----|---------|-------|
| Conditions (A) | 3 | 585.875 | 5.02* |
| Problems (B) | 1 | 50.000 | |
| Sexes (C) | 1 | 990.125 | 8.49* |
| A × B | 3 | 390.917 | 3.35* |
| B × C | 1 | 13.500 | |
| A × C | 3 | 53.208 | |
| A × B × C | 3 | 186.417 | 1.60 |
| Error | 16 | 116.688 | |

* *p* < .05.

irrelevant inquiries. Both groups and individuals were asked to write down all ideas produced so as to prevent groups from having an advantage over individuals of having more time to think, an objection raised by Zagona, Willis, and MacKinnon (1966) to previous studies.

RESULTS

Scoring

For each group of four Ss, whether real or nominal, a single list was made of every response. If any idea was given more than once, the better-worded one was scored.

Individual versus Group Solutions

A factorial analysis of variance (Table 1) shows that the different experimental conditions of problem solving led to significantly different rates in the production of ideas. Based on the means shown in Table 2, the combination of the two mixed conditions (G-I and I-G) produced significantly more solutions ($\bar{X} = 53.5$) than did Condition G ($t = 2.27$; $df = 22$; $p < .05$). Contrary to expectations, however, Condition I resulted in a significantly greater number of ideas than the combination of the mixed conditions ($t = 2.21$; $df = 22$; $p < .05$). Hence, the hypothesis that the number of solutions to a problem should be greater where S works in both a group and an individual setting than where he works in either a group or an individual setting was not supported.

As hypothesized, individual problem solving (I) was more productive in the generation of different solutions than was group problem

TABLE 2

MEAN NUMBER OF IDEAS PRODUCED UNDER EACH CONDITION OF GROUP AND INDIVIDUAL PROBLEM SOLVING

| Condition | Overall mean | Mean of group phase | Corrected ^a mean of group phase | Mean of individual phase | Corrected mean of individual phase |
|-----------|--------------|---------------------|--|--------------------------|------------------------------------|
| I | 63.8 | | | | (63.8) ^b |
| G | 42.9 | | (42.9) ^b | | |
| I-G | 52.6 | 19.7 | 39.4 | 33.0 | 66.0 |
| G-I | 54.5 | 19.0 | 38.0 | 35.5 | 71.0 |

^a Doubled to compensate for half-time.
^b Same as the observed mean.

solving (G) (Table 2: $t = 3.87$; $df = 14$; $p < .01$).

In a similar fashion, within Conditions I-G and G-I the individual sessions produced more solutions than the group sessions (Table 2). Since each part of the mixed conditions lasted 8 min., or half the time for the pure conditions, the mean values of the individual and group phases, taken separately, were doubled. As seen in Table 2, these means parallel closely the means for Conditions I and G.

Other Findings

The source of the significant interaction between problem type and problem solving conditions (Table 1) can be traced to the mixed conditions. As seen in Table 3, for the Tourist problem Condition G-I led to more solutions than Condition I-G; conversely, for the Education problem Condition I-G led to more solutions than Condition G-I. By itself, problem type did not produce any effect.

As also seen in Table 3, males produced consistently significantly more ideas than fe-

males. This difference did not vary from condition to condition or from problem type to problem type, as indicated by the very low F ratios for those interactions involving sex (Table 1).

DISCUSSION

The findings confirm the hypothesis that the individual production of ideas is superior to production within a group setting. This holds true whether individual work precedes, follows, or is independent of group work.

It was also posited that individual and group settings could make their separate and unique contributions to the problem-solving process. However, the hypothesis that mixed conditions should lead to greater productivity than pure conditions was not supported. While the mixed conditions did produce more solutions than the group condition, they produced fewer solutions than the individual condition.

In explanation, it can be seen that the combined mean of both mixed conditions within each problem type is almost at the midpoint between the means for Conditions I and G (Table 3). Since each of the mixed conditions was divided into two equal parts—one group and the other individual—one may argue that the production of ideas is simply a function of the proportion of time spent in an individual situation. In other words, the mixed conditions were superior to the group condition not because they allowed a combination of different working conditions but because they contained a period of individual problem solving. Similarly, since Condition I permitted twice as much time for individual

TABLE 3

MEAN PRODUCTION OF IDEAS UNDER CONDITION, SEX, AND PROBLEM TYPE

| Con- dition | Tourist problem | | | Education problem | | |
|----------------|-----------------|--------|------|-------------------|--------|------|
| | Male | Female | x | Male | Female | x |
| I | 78.0 | 52.0 | 65.0 | 61.5 | 63.5 | 62.5 |
| G | 37.0 | 34.0 | 35.5 | 58.0 | 42.5 | 50.2 |
| I-G | 52.0 | 40.5 | 46.2 | 71.0 | 47.0 | 59.0 |
| G-I | 66.5 | 57.5 | 62.0 | 48.0 | 46.0 | 47.0 |

work as the mixed conditions, its production was greater.

Why do this and other studies (e.g., Dunnette et al., 1963; Taylor et al., 1958) continue to find group brainstorming inferior to working individually? As mentioned earlier, one explanation holds that group discussion channels thinking along similar lines. Another suggests that only those ideas which are socially acceptable will be voiced within a group context.

As an alternative explanation, perhaps group participation has been placed at a disadvantage because only relatively short time intervals, ranging from 5 to 16 min., have been utilized in research. Hence, only the initial spurts in thinking are being tapped. In group sessions most participants might not have time to record their own spurts while listening to and recording those of the others and they may become temporarily forgotten. Allowing longer group sessions might permit the recall of these items and, hence, lead to

the increased production of ideas. It also follows that for group conditions, the production of responses would decrease less over time than the production of ideas under individual conditions.

REFERENCES

- DUNNETTE, M. D., CAMPBELL, J., & JAASTAD, K. The effect of group participation on brainstorming effectiveness for two industrial samples. *Journal of Applied Psychology*, 1963, 47, 30-37.
- MEADOW, A., PARNES, S. J., & REESE, H. Influence of brainstorming instructions and problems sequence on a creative problem-solving test. *Journal of Applied Psychology*, 1959, 43, 413-416.
- OSBORN, A. F. *Applied imagination*. New York: Scribner's, 1953.
- TAYLOR, D. W., BERRY, P. C., & BLOCK, C. H. Does group participation when using brainstorming facilitate or inhibit creative thinking? *Administrative Science Quarterly*, 1958, 3, 23-47.
- ZAGONA, S. V., WILLIS, J. E., & MACKINNON, W. J. Group effectiveness in creative problem-solving tests: An examination of relevant variables. *Journal of Psychology*, 1966, 62, 111-137.

(August 12, 1968)

Manuscripts Accepted for Publication in the
Journal of Applied Psychology

- Processing Auditory Information: Interference from an Irrelevant Cue: J. Richard Simon* and A. M. Small, Jr.: Department of Psychology, University of Iowa, Iowa City, Iowa 52240.
- Impact of Employee Participation in the Development of Pay Incentive Plans: A Field Experiment: Edward E. Lawler, III,* and J. Richard Hackman: Department of Administrative Sciences, Yale University, 2 Hillhouse Avenue, New Haven, Connecticut 06520.
- Consumer Attitudes toward Auto versus Public Transport Alternatives: Frank T. Paine, Allan N. Nash,* Stanley J. Hille, and G. Allen Brunner: Department of Business Administration, University of Maryland, College Park, Maryland 20742.
- Correlates of Employee Evaluations of Pay Increases: J. R. Hinrichs*: IBM Corporation, 112 East Post Road, White Plains, New York 10601.
- Prediction of Job Performance from Assessment Reports: The Use of a Modified Q-Sort Technique to Expand Predictor and Criterion Variance: Garland Y. DeNelsky* and Michael G. McKee: 1020 North Quintana Street, Arlington, Virginia 22205.
- Skimming Lists of Food Ingredients Printed in Different Brightness Contrast: E. C. Poulton*: Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.
- Effect of Interviews on Teacher Selection Decisions: Dale L. Bolton* and Michael E. Hickey: School of Education, Stanford University, Stanford, California 94305.
- A Simplified Method for Increasing Mail-Sorting Efficiency: Thomas Rex Long, Dell Lebo,* and Dorothy Ogden: Child Guidance Clinic, 1635 St. Paul Avenue, Jacksonville, Florida 32202.
- On the Use of Alternative Organisms: Douglas A. Bernstein* and Thomas N. Alloway: University of Illinois, Psychological Clinic, Children's Research Center Building, Champaign, Illinois 61820.
- Relevance of Rater-Ratee Acquaintance in the Validity and Reliability of Ratings: Norman E. Freeberg*: Developmental Research Division, Educational Testing Service, Princeton, New Jersey 08540.
- Effects of Organizational Structure on Correlations between Member Abilities and Group Productivity: Gordon E. O'Brien* and A. G. Owens: Psychology Department, School of Social Sciences, Flinders University, Bedford Park, South Australia 5042.
- Toward an Hypothesis of Work Behavior: Abraham K. Korman*: Department of Psychology, New York University, 21 Washington Place, New York, New York 10003.
- Peer Nominations as Predictors of Short-Term Sales Performance: L. K. Waters* and Carrie Wherry Waters: Department of Psychology, Ohio University, Athens, Ohio 45701.
- Attitudes of Professional Appraisers toward Aesthetics and Appraisal Activities: Jack M. Wright* and J. H. Lemly: American Institutes for Research, P. O. Box 1113, Palo Alto, California 94302.
- Distortion of Drivers' Estimates of Automobile Speed as a Function of Speed Adaptation: Frank Schmidt and Joseph Tiffin*: Department of Psychology, Purdue University, West Lafayette, Indiana 47907.
- Size, Performance, and Potential in Brainstorming Groups: Thomas J. Bouchard, Jr.,* and Melena Hare: Department of Psychology, University of California, Santa Barbara, California 93106.
- Prediction of Driving Behavior Following a Group Driver Improvement Session: Robin S. McBride*: Department of Motor Vehicles, Research and Statistics, P. O. Box 1828, Sacramento, California 95809.
- Use of Time-Lapse Photography in Flight Performance Evaluation: Robert N. Isley* and Paul W. Caro, Jr.: Human Resources Research Office, Division No. 6 (Aviation), P. O. Box 428, Fort Rucker, Alabama 36360.

* Asterisk indicates author for whom address is supplied.

Journal of applied psychology

Table of Contents

| | | |
|--|---|----------|
| Some Results of Three Basic Skills Training Programs in Industrial Setting. | William G. Mollenkopf | 343 |
| Relationship of the Components of an Assessment Center to Management Success. | Herbert B. Wollowick and W. J. McNamara | 348 |
| Vocational Interests of Women: A Factor Analysis of the Women's Form of the SVIB. | Kirk E. Farnsworth | 353 |
| Occupational Group as a Moderator of the Job Satisfaction-Job Performance Relationship. | Richard E. Doll and E. K. Eric Gunderson | 359 |
| Recruiting Door-to-Door Salesmen by Mail. | Valentine Appel and M. R. Feinberg | 362 |
| Use of the Work Components Study with New College-Level Employees. | Robert N. Ford, Edgar F. Borgatta, and George W. Bohrnstedt | 367 |
| Influence of Sex Roles on the Manifestation of Leadership. | Edwin I. Megargee | 377 |
| The GRE Psychology Test as an Unobtrusive Measure of Motivation. | Robert B. Ewen | 383 |
| Correlates of Job Satisfaction and Job Dissatisfaction among Female Clerical Workers. | L. K. Waters and Carrie Wherry Waters | 388 |
| Creativity and Academic Major: Business versus English Majors. | Russell Eisenman | 392 |
| Forced-Choice Preferences and Response Latency. | Bruce O. Bergum and Roger P. Dooley | 396 |
| A Validation Study of Polygraph Examiner Judgments. | Philip J. Bersh | 399 |
| MMPI Scores as Related to Age, Education, and Intelligence among Male Job Applicants. | Fred J. Thumin | 404 |
| The SVIB for Women and Demographic Variables in the Prediction of Occupational Tenure. | Thomas K. Stone and Gary T. Athelstan | 408 |
| Feedback and Response Mode in Performing a Bayesian Decision Task. | David W. Martin and Charles F. Gettys | 413 |
| Interference between Concurrent Tasks of Driving and Telephoning. | I. D. Brown, A. H. Tickner, and D. C. V. Simmonds | 419 |
| Comparison of "Real Life" Assessments of Management Potential with Situational Exercises, Paper-and-Pencil Ability Tests, and Personality Inventories. | J. R. Hinrichs | 425 |
| Processing Auditory Information: Interference from an Irrelevant Cue. | J. Richard Simon and A. M. Small, Jr. | 433 |
| Image of Industrial Psychology among Personnel Administrators. | George C. Thornton, III | 436 |
| List of Manuscripts Accepted. | | 398, 438 |

KENNETH E. CLARK
University of Rochester

Consulting

GEORGE E. BRIGGS
Ohio State University

MARVIN D. DUNNETTE
University of Minnesota

NORMAN
Educational Testing Service

LEONARD D. GOODSTEIN
University of Cincinnati

EDWIN R. HENRY
University of Rochester

JOHN HOLLAND
American College Testing
Program

CLIFFORD E. JURGENSEN
Minneapolis Gas Company

DONALD G. MARQUIS
Massachusetts Institute of
Technology

LAURENCE S. MCGAUGHRAN
University of Houston

QUINN MCNEMAR
University of Texas

WARREN T. NORMAN
University of Michigan

HAROLD F. ROTHE
Beloit Corporation

THOMAS A. RYAN
Cornell University

ALEXANDER G. WESMAN
Psychological Corporation

CLARK L. WILSON
Instituto Centroamericano de
Administracion de Empresas

This journal gives primary consideration to original investigations in any field of applied psychology except clinical psychology, although a descriptive or theoretical article may be accepted if it represents a special contribution in the applied field. Quantitative investigations of interest or value to psychologists working in the following broad fields will be considered: personnel research in business, industry, and government; engineering psychology; industrial working conditions; research on opinion and morale factors; job analysis and classification research; market and advertising research; vocational and educational prognosis, diagnosis, and guidance at the secondary and college level.

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or they may be obtained from the Editor or from the APA Central Office.

Manuscripts should be addressed to the Editor:

Dr. Kenneth E. Clark, Dean
College of Arts and Science
302 Morey Hall
University of Rochester
Rochester, N. Y. 14627

Authors are strongly advised to follow the general directions given in the *Publication Manual of the American Psychological Association* (1967 Revision). Special attention should be given to the section on the preparation of the references (p. 39). In reference lists, give journal titles in full; do not abbreviate. *All copy must be double spaced, including the references.* All manuscripts should be submitted in *duplicate*, one of which should be an original typed copy; author's name should appear only on title page. Dittoed and mimeographed copies are not acceptable and will not be considered. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail and to check carefully the typing of the final copy.

The following policies govern reprinting or other forms of reproduction of APA-copyrighted journal materials. Permission by the American Psychological Association must be requested for reprinting all tables, figures, and text of more than 500 words (total from one article) in length. There will be a charge of \$5.00 per page for such permission except when the requester is the author of the article to be reprinted, or the article is to be reproduced in a limited number of copies solely for instructional purposes.

Published bimonthly by the
American Psychological Association
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W.
Washington, D. C. 20036

\$10.00 per volume

\$2.00 per issue

HELEN ORR
Managing Editor

CLAIRE PILLIOD
Technical Editor

HENRY S. SALMON
Advertising and
Business Manager

ROBERT D. GORMAN
Subscription Manager

Subscriptions, orders, and business communications should be addressed to the American Psychological Association, 1200 Seventeenth St., N. W., Washington, D. C. 20036. Address changes must reach the subscription office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Second-class postage paid at Lancaster, Pa.

SOME RESULTS OF THREE BASIC SKILLS TRAINING PROGRAMS IN AN INDUSTRIAL SETTING

WILLIAM G. MOLLENKOPF¹

The Procter & Gamble Company

Test scores obtained before and after instruction were analyzed to evaluate outcomes of three types of training. The training programs were designed to improve basic skills of present and prospective employees in production, office, and laboratory work. In most of the learning situations, groups made significant gains. Individual true gains also were studied, using a method developed by Lord.

In recent years, a number of industrial firms have offered to employees or prospective employees new training programs intended to improve basic skills such as reading and arithmetic (Burck, 1968; Gassler, 1967; Gustaitis, 1967; Janger, 1969). Such programs have particular relevance for those who lack adequate opportunity for good education.

The following reports describe three types of training programs, carried out in Procter & Gamble in 1967-68, oriented toward (a) production jobs, (b) typing and secretarial work in offices, and (c) technician positions in laboratories. Following description of the training, test data are presented which reflect in part the outcomes of participation in the program. These group data are supplemented by information about individual gains, presented in a later section.

PRODUCTION EMPLOYEE GROUPS

The MIND (Methods of Intellectual Development) program (Gustaitis, 1967) was used with groups at two manufacturing plants. This program consisted of two parts, Communications Skills and Mathematics. The first part included instruction on vocabulary,

word building with prefixes and suffixes, reading aloud, spelling, and grammar. This instruction was given largely through the use of workbooks. The Mathematics training was given principally through the use of tapes which presented problems of increasing complexity at faster and faster rates. Both parts provided some opportunity for adaptation to individual differences. Instruction was given before or after the work shift, that is, on the employee's own time. Costs of instruction were borne by the employer.

Plant A

At this plant, 17 men participated in part or all of the MIND program. Of those who continued through the program, 10 took both Communications Skills and Mathematics, 4 took Mathematics only, and 1 took Communications Skills only. Two dropped out during the course. Ages of the participants ranged from 21 to 47, with a median of 29. Years of formal schooling reported varied from 6 to 14; the median was 12. About one-third of the men were Negroes.

Instruction was given in two 2-hr. sessions per week over a 20-wk. period, on a total of 78 days. Trainees, other than those who dropped out, varied in number of days attended, from a high of 72 to a low of 33,

¹Requests for reprints should be sent to the author, The Procter & Gamble Company, P. O. Box 599, Cincinnati, Ohio 45201.

with a median of 66. Shift rotation was a frequent cause of nonattendance.

Parallel forms of the Paragraph Meaning and Arithmetic Computation sections of the Stanford Achievement Test, Intermediate II Battery (Kelley, Madden, Gardner, & Rudman, 1964) were given prior to and following completion of the MIND program to provide information on changes in skill levels as a result of the training.

Paragraph meaning. Preinstruction grade-equivalent scores for the 11 men who took the Communications Skills training ranged from 5.0 to 12.9, with a mean of 8.5 and a median of 8.0. After-instruction scores ranged from 5.6 to 11.5, with a mean of 9.4 and a median of 10.6. The difference between the means did not reach significance at the .05 level ($t = 1.33$, $df = 9$).

Arithmetic computation. Grade-equivalent scores obtained prior to the beginning of instruction by the 14 men who participated in training in this area ranged from 6.0 to 11.7, with a mean of 8.2 and a median of 7.9. At the end of instruction, scores ranged from 10.5 to 12.9, with a mean of 12.1 and a median of 12.6. More than half of the group had a postcourse score at or very close to the maximum grade score obtainable (12.9), so there was a definite ceiling effect here. The difference between means was significant beyond the .01 level ($t = 3.27$, $df = 11$).

Plant B

At this plant, 29 men participated in the program. Their ages ranged from 20 to 52, with a median of 37. Years of school reported as completed ranged from 5 through 12, with a median of 10. About two-thirds of the men were Negroes.

Instruction was offered in three 2-hr. sessions per week, over a 20-wk. period. Each participant could have received 116 hours of instruction had he attended all sessions. However, actual hours for the 27 men for whom this information was available ranged from a low of 14 to a high of 94. Nine persons dropped out; 1 man started late.

At this plant, parallel forms of the Reading and Arithmetic Computation Tests of ABLE (Adult Basic Learning Examination; Karlson, Madden, & Gardner, 1967) were given

before and at the end of the MIND program. Besides the nine dropouts and the man who began late, there were three other persons for whom test records were incomplete.

Reading. Preinstruction grade-equivalent scores ranged from 4.1 to 9.0+ (the top score obtainable), with a mean of 7.4 and a median of 8.1. Two were at the top score. Scores at the end of instruction ranged from 4.7 to 9.0+, with a mean of 8.0 and a median of 8.7. Six of the group attained the top score; again, a definite ceiling effect existed here. The difference between means was significant at the .05 level ($t = 2.85$, $df = 14$).

Arithmetic computation. Preinstruction grade-equivalent scores ranged from 3.9 to 6.9, with a mean of 5.2 and a median of 5.1, scores clearly and appreciably lower than those for Reading. Scores after instruction ranged from 3.0 to 9.0+, with a mean of 6.8 and a median of 7.3. Four men achieved the top score. The difference between means was significant at the .05 level ($t = 2.88$, $df = 14$).

OFFICE WORKER GROUPS

During the 2 years, four small groups of women were given special training in order that they might be better prepared for secretarial and clerical positions in the main offices and technical centers of the firm. All participants were Negroes. Most were recent high school graduates, and single. A total of 43 women began the training, and all but 2 completed it. Most of these were offered, and accepted, regular office positions.

Among the instructional topics, four content areas were given principal attention: spelling, grammar, vocabulary, and arithmetic. In spelling, emphasis was placed on words frequently misspelled in office work. Numerous informal oral quizzes were used, and students were encouraged to read broadly and use the dictionary to build their vocabularies.

For instruction in grammar, a programmed text was used with three groups, but abandoned as not sufficiently suitable for the fourth. With the fourth group, emphasis was placed on conversational grammar.

For improving arithmetic skills, large numbers of problems involving basic operations were given, and difficulty and time pressure

were gradually increased. Word problems were introduced to give practice in thoughtful application of number skills.

Arrangements for the instruction varied somewhat from group to group, but typically involved two or three sessions per week over a period of several weeks, with a total of about 60 hr. Participants were paid as part-time employees.

Pre- and posttest scores were available for most participants completing the program from administration of two parallel forms of a five-part test devised by the firm's personnel research group. Areas covered by this test were spelling, expression (grammar), filing, arithmetic, and reasoning.

Spelling. The test used presents a list of 30 frequently misspelled words, about half of which are spelled correctly. The task is to indicate whether the word is spelled correctly or not. Pretest scores had a mean of 20.6 and a median of 20.0; posttest scores had a mean of 22.7 and a median of 24.0. The difference between the pre- and post-instruction means was significant at the .01 level ($t = 3.46$, $df = 38$).

Expression. The 15 items in this test involve incomplete sentences. The task is to choose words or punctuation marks which best complete the sentences in terms of good usage. The mean of the before-training scores was 8.6, and the median was 8.5. For the after-training scores, the mean was 9.2 and the median 9.0. The difference between means was significant at the .05 level ($t = 2.30$, $df = 39$).

Filing. Fifty names of firms are given, and the task is to "file" these in the appropriate simulated file pockets, working as rapidly and accurately as possible. The before-training scores had a mean of 28.6 and a median of 29.0; the mean of the after-training scores was 34.8 and the median was 36.0. The difference between means was significant at the .01 level ($t = 3.82$, $df = 39$).

Arithmetic. The test used consists of 60 short problems in addition, subtraction, multiplication, and division. Answers are provided and are to be marked as right or wrong. The mean of the before-training scores was 25.2 and the median, 26.0. For after-training scores, the mean was 30.3 and the median

was 30.5. The difference between means was significant at the .01 level ($t = 4.81$, $df = 39$).

Reasoning. This test consists of 10 verbal analogies items and 10 arithmetic reasoning items. Before-training scores had a mean of 8.1 and a median of 8.0. For after-training scores, the mean was 10.0 and the median was 10.0. The difference between means was significant at the .01 level ($t = 3.75$, $df = 39$).

LABORATORY TECHNICIAN TRAINING GROUP

This program was designed to offer training to persons potentially qualifiable for employment as laboratory technicians, but lacking the present levels of competence ordinarily sought in areas such as elementary mathematics, basic chemistry, and English usage.

Twenty-six individuals began the training program. Nineteen were male, 7 female. All but one were high school graduates or the equivalent. Most of the participants were Negroes.

Instruction in mathematics and English placed strong emphasis on use of workbooks. The two instructors also utilized many other resources: books on communication skills and general mathematics; dictionaries; periodicals such as *Science News*, *Readers' Digest* and the *New York Times*; audiovisual materials and equipment; and commercially developed programs concerned with improvement of reading, listening, and study skills.

After an initial orientation period, participants spent approximately half their time in the classroom activities described above and the other half in on-the-job training in laboratory situations, learning more about the work of a technician. Class participants were employed with the understanding that satisfactory completion of this classroom-job-training-program would lead to regular employment as a technician trainee.

Twenty-five of the classroom participants completed the formal instructional program, and all of these were offered regular employment as technician trainees.

Learning measures. Throughout the course of the training period quizzes were frequently used. However, in only one instance was

TABLE 1
INDIVIDUAL GAINS

| Study group | Test | Estimated true individual gains | | |
|--------------|---|---------------------------------|--------|-------------------------|
| | | Range | Median | Percent showing + gains |
| Plant A | Paragraph Meaning (grade equivalent score) | -1.5 to +5.2 | + .5 | 73 |
| | Arithmetic Computation (grade-equivalent score) | +1.7 to +5.1 | +4.2 | 100 |
| Plant B | Arithmetic Computation (grade-equivalent score) | -.9 to +3.0 | +1.8 | 81 |
| Office group | Spelling (raw score) | -1.6 to +6.2 | +2.5 | 87 |
| | Filing (raw score) | -11.8 to +22.2 | +7.2 | 90 |
| | Numerical Operations (raw score) | -2.7 to +12.1 | +5.0 | 98 |
| | Reasoning (raw score) | -.3 to +5.0 | +1.7 | 95 |

there a before-and-after type of measurement.

During the orientation period, the students were given a 100-item test on arithmetic processes and manipulation of whole numbers, fractions, decimals, percentages, ratios and proportions, and formulas. Scores ranged from 36 to 90, with a mean of 61.5. Seven weeks later, the same test was administered again without prior notice. The nature of the content was such that in the intervening time many similar problems would have been encountered, so recall of the original answers would be quite unlikely. For every one of the 25 students taking the test again, an increase in score occurred, ranging from half a point (by one of the top scorers on the first administration) to 29½ points (by the lowest scorer on the first attempt). Final scores varied from 45 to 98, with a mean of 75.2. The difference between means was significant at the .01 level ($t = 4.28$, $df = 24$).

INDIVIDUAL GAINS

In addition to changes in mean scores, it is of interest to assess the gains made by individuals during the course of instruction. Often the observed change in test score is treated as a gain. But Lord (1956, 1958, 1963) and others have pointed out that the use of the observed score difference as a measure of gain may be quite misleading. Both regression effects and errors of measurement need to be taken into account.

For several of the sets of plant and office data, it was possible to apply the regression

equation approach developed by Lord (1956) which assumes the standard errors of measurement are the same for the pre- and post-test scores. Table 1 summarizes the results obtained for those situations in which this approach fit the data well enough to yield meaningful outcomes.

For the first three rows of this table, gains are stated in terms of grade-equivalent scores which give the reader a feeling of the magnitude of the individual true gains estimated. For the office group, it may help to note that the median gain reported was more than half the magnitude of the standard deviation of the pretest scores in each instance.

DISCUSSION

For each of the three programs—production employee, office worker, and technician trainee—the test results for the group or groups involved showed an increase in average score over the training period with the change in mean significant at or beyond the .05 level in all but one instance, that for the smallest group.

Methods of instruction varied from one program to another. With the differences in content and students involved, comparison of the relative effectiveness of the instructional approaches is not possible. What can be said is that each instructional method was associated with improvement in the group average in each instance.

Considering that the production workers and office trainees typically spent well under

100 hr. in the training program, and that each program had at least two major emphases, the gains made seem impressive. Results presented here are evidence that workers, some well along in years, can make sizable gains in basic skills such as arithmetic. Unfortunately, relatively little good data are as yet available on the outcome of such special training programs. So the gains observed in the situations reported on here cannot be measured against reasonable expectations for outcomes from part-time instructional programs of several weeks to several months duration. Such expectations can be built up only from a number of reports from various sources. There is a strong need for more data of the type reported here.

REFERENCES

- BURCK, G. A new business for business: Reclaiming human resources. *Fortune*, 1968, 77, 159-161, 198.
- GASSLER, L. S. How companies are helping the uneducated worker. *Personnel*, 1967, 44, 47-55.
- GUSTAITIS, R. Private industry's factory classrooms. *The Reporter*, 1967, 37, 23-24.
- JANGER, A. R. "New start"—for the harder hard-core. *The Conference Board Record*, 1969, 6, 10-19.
- KARLSEN, B., MADDEN, R., & GARDNER, E. F. *ABLE handbook*. New York: Harcourt, Brace & World, 1967.
- KELLEY, T. L., MADDEN, R., GARDNER, E. F., & RUDMAN, H. C. *Stanford Achievement Test: Directions for administering*. New York: Harcourt, Brace & World, 1964.
- LORD, F. M. The measurement of growth. *Educational and Psychological Measurement*, 1956, 16, 421-437.
- LORD, F. M. Further problems in the measurement of growth. *Educational and Psychological Measurement*, 1958, 18, 437-454.
- LORD, F. M. Elementary models for measuring change. In C. B. Harris (Ed.), *Problems in measuring change*. Madison, Wisc.: University of Wisconsin Press, 1963.

(Early publication received April 17, 1969)

RELATIONSHIP OF THE COMPONENTS OF AN ASSESSMENT CENTER TO MANAGEMENT SUCCESS

HERBERT B. WOLLOWICK AND W. J. McNAMARA¹

IBM, Armonk, New York

The purpose of this study was to determine the validity of an assessment-center approach in predicting management potential and to determine the relative value of the components of the program. Results indicate that the approach is valid and that situational tests add to the predictiveness of paper-and-pencil tests. Also demonstrated was greater predictiveness through statistical combination of the program variables, rather than a subjectively derived overall rating.

Because of the ever-increasing need for capable managers, many companies are taking a critical look at their traditional methods of selecting individuals for managerial assignments. Further, there is a heightened interest in identifying the potential manager early in his career so that he may be properly groomed for high-level managerial responsibilities. The techniques developed by the Standard Oil Company of New Jersey (Laurent, 1961) clearly demonstrate the possibilities of early identification of managerial talent utilizing paper-and-pencil tests and inventories.

Another approach, which involves the use of situational tests (Flanagan, 1954) in addition to written objective tests, has been adopted recently by several business organizations. The most extensive use of this type of assessment procedure for identifying managerial talent has been by AT&T (Bray & Grant, 1966). Merits of this method have also been demonstrated by Albrecht, Glaser, and Marks (1964).

One purpose of this paper is to present additional data on the validity of the assessment-center technique based on its use in a large electronics concern. The data available, however, are extensive enough to make possible the investigation of other crucial questions relative to the assessment center. One question frequently raised concerns the advantages of the situational tests utilized in the assessment-center technique over and above the use of written (paper-and-pencil) instruments alone. Stated in another way, the

question is how much additional information can be gained by the use of group and individual situational tests. These are the most costly aspects of the assessment-center approach. If additional validity is obtained from the situational exercises, is the additional gain sufficiently large to justify the time and cost involved?

Another question relates to the length and complexity of the assessment-center approach. How many tests are necessary, how many individual exercises should be utilized, and how many group exercises should be a part of the program? A further question relates to the emphasis that should be placed on the various elements of the program. In the final evaluation of an individual participant in an assessment-center program, how much weight should be given to the written tests, how much weight should be given to the situational exercises, and more specifically, how much weight should be given to each of the individual written tests and situational exercises in the assessment program?

METHOD

The Ss in the study were 94 men from two divisions of a large electronics firm. They were all from lower and middle management positions and participated in the assessment program between September, 1964, and June, 1965. The men selected to participate in the program had been designated as having above-average potential for advancement; they were not a random selection of men at their level in the business. Most of the men were in their late 20s or early 30s, and their educational level was approximately 2 yr. beyond high school.

The criterion used was the increase in managerial responsibility as of January, 1968, approximately 3 yr. after participation in the assessment program.

¹ Requests for reprints should be sent to Walter J. McNamara, IBM Corporation, Armonk, New York 10504.

The measure of managerial responsibility used was position-code level, a two-digit number that is determined by a job analysis of the position, taking into account the number of people supervised, complexity of the job, financial responsibility involved, skill required, etc. Thus, two individuals at the same level of management could have different position-code levels depending on the managerial responsibility involved in their job. This criterion was used rather than merely the level of management attained, because it is a finer measure of advancement. Thus, in this study a scale of 12 increments or management steps was involved.

The two-day assessment program was very similar to that described by Bray and Grant (1966), and the group situational exercises were identical to those described by Greenwood and McNamara (1967).

The written tests used in the assessment program included cognitive ability tests, personality inventories, measures of leadership ability, and background history. The specific tests and subscores were (a) Gordon Personal Profile—four scores: Ascendancy, Responsibility, Emotional Stability, and Sociability. (b) Gordon Personal Inventory—four scores: Cautiousness, Original Thinking, Personal Relations, and Vigor. (c) Fleishman's Leadership Questionnaire (LOQ)—two scores: Structure and Consideration. (d) Background and Contemporary Data Form (BCD)—an in-company developed biographical inventory scored with two keys: one for self-confidence and one for characteristics of a successful manager. (e) Otis Employment Test—total score. (f) School and College Ability Test—total score combining verbal and arithmetic sections.

The group exercises were

1. **Leaderless Group Discussion:** Each participant is required to make a 5-min. oral presentation of a candidate for promotion and then subsequently defend his candidate in a group discussion with five other participants. Characteristics rated are aggressiveness, persuasiveness or selling ability, oral communications, self-confidence, resistance to stress, energy level, and interpersonal contact.

2. **Manufacturing:** Six participants are required to work together as a group and operate a manufacturing company in an effective manner. They are required to purchase material, manufacture a product, and sell it back to the market. Included in the exercise is a product forecast, specific prices for raw materials, and completed products which fluctuate during the exercise. The participants are not given assigned roles. Both verbal and physical activity are involved. Characteristics rated are aggressiveness, persuasiveness or selling ability, resistance to stress, energy level, interpersonal contact, administrative ability, and risk-taking.

3. **Task Force:** Participants are given data from which they are to form a group recommendation to the president of a company regarding which of three alternative courses of action he should take for expansion of the company's activities. Characteristics rated are aggressiveness, persuasiveness or sell-

ing ability, oral communications, self-confidence, energy level, decision-making, and interpersonal contact.

The individual exercises included

1. **In-Basket:** Each individual is required to assume the position of a personnel manager in a manufacturing plant for one hour and a half. In this time, he must deal with internal and external correspondence which has accumulated since his predecessor was assigned to another position. For about 30 min., the participants are individually interviewed by an observer regarding the actions that they have taken and the reasons for their decisions. Characteristics rated are oral communications, planning and organizing, self-confidence, written communications, decision making, risk taking, and administrative ability.

2. **Job Environment Report:** This exercise requires the individual to describe in writing his job, the things he likes about his work, the things he dislikes, and his relationship to his peers, subordinates, and supervisors. The participant is allowed 1 hr. to complete this exercise. Characteristics rated are planning and organizing, self-confidence, written communications, and interpersonal contact.

3. **Stock Market:** The participant is given the role of an investor. He buys and sells stocks from a set money reserve under conditions of changing market values. The individuals write their own buy and sell orders, calculating their expenditures and profits. Characteristics rated are planning and organizing, self-confidence, resistance to stress, decision making, administrative ability, and risk-taking.

In addition to the scores on the written tests, several ratings for each participant in the program were obtained—

1. **Overall Rating (OAR):** At the conclusion of the 2-day assessment program, the four observers who formed the assessment staff and who were all operating management personnel at least two levels above the participants, discussed each participant's performance and assigned him an OAR, taking into consideration all of the variables in the program. The OAR was on a 5-point scale: (a) exceptional potential for advancement, (b) above-average potential, (c) average potential, (d) below-average potential, and (e) no potential.

2. **Exercise Ratings:** In the group's exercises, each of the three observers rated the six participants on a 5-point scale: (a) much more effective than most of the group, (b) somewhat more effective, (c) about as effective as most of the group, (d) relatively ineffective, and (e) ineffective. The three observer ratings in each group exercise were averaged to produce the participant's rating for the exercise. For the individual exercises a participant's rating was based on a single observer. Reliability of the observer ratings in this program has been described in a previous paper by Greenwood and McNamara (1967).

3. **Characteristic Ratings:** 12 characteristics were designated as important elements of managerial performance: Self-Confidence, Written Com-

munications, Administrative Ability, Interpersonal Contact, Energy Level, Decision Making, Resistance to Stress, Planning and Organizing, Persuasiveness or Selling Ability, Aggressiveness, Risk Taking, and Oral Communications.

The specific characteristics (usually 4–7) that were displayed in a particular exercise (group or individual) were rated on a 5-point scale by a single observer. Then the ratings from the several exercises were averaged to obtain a final characteristic score.

RESULTS

In Table 1 are the correlations of the 33 variables (OAR, 14 test scores, 6 exercise

ratings, and 12 characteristic ratings) with change in position level. From this table it will be noted that the OAR correlates .37 with the criterion (significant at the .01 level). However, one of the test scores, GPP, Ascendancy, correlates .39 with the criterion.

As a matter of fact, three of the eight scales of the Gordon tests are significantly related to the criterion. However, neither the mental ability tests nor the LOQ are significantly related to the criterion measure.

Two of the three group exercises (Leaderless and Manufacturing) provided significant correlations, in addition to one of the individual exercises (In-Basket).

Of the 12 characteristics, 9 had significant correlations with the criterion. From a factor analysis of the characteristics measured in the assessment program, McNamara and Greenwood (1967) identified 5 characteristics as belonging to an activity dimension (Self-Confidence, Energy Level, Persuasiveness, Aggressiveness, and Oral Communications). It will be noted that all 5 of these characteristics are significantly related to the criterion. Of the test scores included in the program, it will also be noted that Ascendancy, Vigor, and Self-Confidence from the BCD are all significantly related to the criterion. Thus, it is apparent that for this population, an activity factor, however measured, is significantly related to success in management.

Table 2 shows multiple *R*s for each type of predictor (i.e., tests, exercises, characteristics). Multiple *R*s are also shown for all possible combinations of tests, exercises, and characteristics. For each multiple *R*, the percentage of variance (*R*²) is also given. The figures shown are cut off at the point where the next variable did not add significantly to the correlation (*F* test each term). The table shows a multiple *R* of .39 for exercises alone, which accounts for 15% of the criterion variance and a multiple *R* of .45 for tests alone, which accounts for 20% of the criterion variance. Combining tests with exercises and characteristics increases the multiple *R* to .62, which accounts for 38% of the criterion variance. In comparison, the OAR correlated .37 with the criterion, which accounts for 14% of the criterion variance.

TABLE 1
CORRELATIONS WITH CHANGE IN POSITION LEVEL
(*N* = 94)

| Variable | <i>R</i> |
|-------------------------|----------|
| Tests | |
| GPP | |
| Ascendancy | .39** |
| Responsibility | -.12 |
| Emotional Stability | -.18 |
| Sociability | .23* |
| GPI | |
| Cautiousness | -.05 |
| Original Thinking | .05 |
| Personal Relations | -.17 |
| Vigor | .32** |
| LOQ | |
| Structure | .13 |
| Consideration | -.11 |
| SCAT—Total | .11 |
| Otis—Total | .07 |
| BCD | |
| Success | .14 |
| Self-Confidence | .23* |
| Exercises | |
| Manufacturing | .28** |
| Leaderless | .25** |
| Task Force | .15 |
| In-Basket | .32** |
| Job Environment | .07 |
| Stock Market | -.07 |
| Characteristics | |
| Self-Confidence | .32** |
| Written Communications | .29** |
| Administrative Ability | .02 |
| Interpersonal Contact | .00 |
| Energy Level | .26** |
| Decision Making | .29** |
| Resistance to Stress | .26** |
| Planning and Organizing | .23* |
| Persuasiveness | .22* |
| Aggressiveness | .24* |
| Risk Taking | .11 |
| Oral Communications | .22* |
| Overall rating | .37** |

* *p* < .05 (.20 and above).
** *p* < .01 (.25 and above).

TABLE 2

MULTIPLE R s AND R^2 FOR THE THREE TYPES OF PREDICTORS SEPARATELY AND COMBINED

| Predictors | R | R^2 |
|-------------------------------------|-----|-------|
| Tests | .45 | 20 |
| Exercises | .39 | 15 |
| Characteristics | .41 | 23 |
| Tests & exercises | .54 | 29 |
| Characteristics & exercises | .52 | 27 |
| Characteristics & tests | .55 | 30 |
| Tests, characteristics, & exercises | .62 | 38 |

Table 3 details the variables that contribute significantly to each multiple R in stepwise order. It is particularly interesting to note the order of the variables in the multiple R for all three types of measurement combined. First there is a test variable, then an exercise variable, then a characteristic variable, a test variable, an exercise variable, a characteristic variable, and then finally a test variable. This sequence indicates that all three types of variables contribute heavily to the predictive success of the program.

In addition, it will be noted that Ascendancy, Vigor, and the Manufacturing Rating are activity measures. Administrative Ability and Interpersonal Contact have low correlations with the criterion but fairly high correlations with one or more of the activity measures. Additionally, the beta weights for Administrative Ability and Interpersonal Contact are negative, indicating that they are operating as suppressor variables. Furthermore, if that portion of the In-Basket rating attributable to Administrative Ability is subtracted, the activity dimension of the interview accounts for the remaining variance.

Bray and Grant (1966) present predictive correlations (against salary progress) of about the same magnitude for similar characteristics and exercises. However, their ability measures (SCAT and others) yielded larger correlations, and their personality measures (EPPS and Guilford and Martin) yielded smaller correlations. It is difficult to determine why these differences exist. One possible explanation for the greater predictiveness of the ability measures in the AT&T study is that the group might have been more heterogeneous. The smaller correlations for the

personality measures in the AT&T study may be attributed to the different tests used or to differences in the organization involved.

DISCUSSION

It is evident from the above results that the subjectively derived OAR utilized in this assessment program is a valid predictor of management success. The correlation of .37 (significant at .01 level) between the OAR and change in the position-level criterion substantiates this conclusion. However, the data indicates that much higher validities may be achieved by an empirical combination of the scores and ratings ob-

TABLE 3
VARIABLES CONTRIBUTING TO MULTIPLE R s
IN STEPWISE ORDER

| Variable | R |
|-------------------------------------|-----|
| Tests | |
| Ascendancy | .39 |
| Vigor | .45 |
| Exercises | |
| In-Basket | .32 |
| Manufacturing | .39 |
| Characteristics | |
| Self-Confidence | .32 |
| Written Communications | .38 |
| Administrative Ability | .41 |
| Tests & exercises | |
| Ascendancy | .39 |
| In-Basket | .46 |
| Vigor | .51 |
| BCD Success Key | .52 |
| LOQ Consideration | .54 |
| Characteristics & exercises | |
| Self-Confidence | .32 |
| In-Basket | .38 |
| Administrative Ability | .44 |
| Manufacturing | .49 |
| Interpersonal Contact | .52 |
| Characteristics & tests | |
| Ascendancy | .39 |
| Written Communications | .46 |
| Vigor | .50 |
| Administrative Ability | .52 |
| Self-Confidence | .55 |
| Tests, characteristics, & exercises | |
| Ascendancy | .39 |
| In-Basket | .46 |
| Administrative Ability | .51 |
| Vigor | .56 |
| Manufacturing Rating | .59 |
| Interpersonal Contact | .60 |
| BCD Success Key | .62 |

tained in the program. Table 2 indicates that multiple R s—based on tests alone, characteristics alone, or exercises alone—give higher validities than their subjective combination.

Additionally, if all three types of measurement are combined in a multiple R , the resulting correlation is further substantially increased. This indicates that all three measures contribute materially to the validity of the overall program. Table 3 shows the variables that contribute significantly when all three types of measurement are combined. As indicated in the results section, all three types of variables (tests, exercises, and characteristics) are equally represented in this combination multiple.

An important question frequently raised (Bray & Grant, 1966) can also be partially answered from these results. That is, can the assessment procedure be justified in light of its additional cost and time compared to the use of paper-and-pencil tests alone? The multiple R for tests alone (Table 2) is .45 with an R^2 of .20. The multiple R for tests, characteristics, and exercises combined is .62 with an R^2 of .38. Inclusion of the elements unique to the assessment-center procedure, therefore, nearly doubles the criterion variance accounted for. This indicates that the assessment procedure makes a substantial unique contribution to the prediction of management success.

Bray and Grant (1966) arrived at a similar conclusion by a more indirect method. They partialled out mental ability (SCAT) from judged ability and found that reliable variance still remained. They concluded that "the results thus indicate that the assessment process does contribute more than can be gained by the simple administration of paper-and-pencil ability measures." The results from this study certainly support and confirm this conclusion.

The results obtained may also be valuable in the future development of a shorter assessment program with maximum predictive results. When all of the variables are combined in a multiple R , only 7 of the 32 variables studied contributed significantly. Based on these results, it may be possible to consider eliminating the paper-and-pencil tests not

contributing to the predictive validity and substituting other tests. However, eliminating any of the exercises may not give enough opportunity to establish usable characteristic ratings, which are presently based on all the exercises. For example, utilization of only three of the six exercises may significantly reduce the reliability and validity of our characteristic ratings. More research is thus needed to determine which exercises could best be eliminated without an adverse effect on the characteristic ratings. Some preliminary work has already been done on the substitution of another exercise for the Task Force Exercise.

Another interesting possibility for improvement of the program is indicated by the results. The subjectively derived combination of the variables (OAR) correlated .37 with the criterion, while the statistical combination gave a multiple of .62. This suggests that instead of deriving an OAR by subjective mean, it might be done more profitably by a statistical procedure. This should greatly increase the predictiveness of the program. In order to accomplish this, however, further research and cross-validation of these results is needed. Meanwhile, improvement is possible by emphasizing in the subjective determination of the OAR those variables that show the highest correlation with the criterion.

REFERENCES

- ALBRECHT, P. A., GLASER, E. M., & MARKS, J. Validation of a multiple-assessment procedure for managerial personnel. *Journal of Applied Psychology*, 1964, 48, 351-360.
- BRAY, D. W., & GRANT, D. L. The assessment center in the measurement of potential for business management. *Psychological Monographs*, 1966, 80 (17, Whole No. 625).
- FLANAGAN, S. C. Some considerations in the development of situational tests. *Personnel Psychology*, 1954, 7, 461-464.
- GREENWOOD, J. M., & McNAMARA, W. J. Interrater reliability in situational tests. *Journal of Applied Psychology*, 1967, 51, 101-106.
- LAURENT, H. A. Summary of early identification of management potential. Research Project, 1961, Social Science Research Division of the Employee Relations Department, Standard Oil Company.
- McNAMARA, W. J., & GREENWOOD, J. M. Evaluation of the personnel assessment program. Personnel Research Studies, IBM, 1967.

(Early publication received January 22, 1969)

VOCATIONAL INTERESTS OF WOMEN: A FACTOR ANALYSIS OF THE WOMEN'S FORM OF THE SVIB¹

KIRK E. FARNSWORTH²

Iowa State University

This study presents an attempt at better understanding of the women's form of the Strong Vocational Interest Blank (SVIB). The responses of 671 women-in-general to each of the 400 items on Form T400R were intercorrelated and factor analyzed using the multiple group method and a variation of Lawley's maximum likelihood procedure. Representative items for each of the resulting 9 common and 17 more specific factors are noted, and descriptive labels are suggested. The procedure and resulting information are seen as distinctly different from previous analyses.

It has been estimated that by 1970 one out of every three workers will be a woman (Miller, 1964); by 1980 75% of all college-educated women will be at work (Keyserling, 1965). Accordingly, Arns (1958), Miller (1964), and others have emphasized the crucial importance of adequate vocational and educational guidance for females. Since the measurement of interests comprises an important segment of such guidance, what is sorely needed is more precise knowledge about the psychological tests that are used to measure the vocational interests of women.

Construction of women's vocational interest scales is difficult. A main cause is the inability to obtain homogeneous criterion groups, which results in difficulty in measuring heterogeneous interest patterns among women (Super & Crites, 1962). Hogg (1928) explained the measurement problems on the ground that women work not for the love of work but just to be busy, or that they tend to choose the occupations which offer the least resistance. Strang (1941) also concluded that women enter certain occupations for reasons other than genuine interest in the job. Strong (1943) specified the reasons as convenience and "stopgap" until marriage—referred to by others as premarital, noncareer,

filler, something-to-fall-back-on-in-case, and occupational insurance.

According to Super (1945), women's interests are more nearly universal (nonspecific) than men's. Similarly, Darley and Hagenah (1955, p. 70) claim that "women's interests are generally less channelized or less professionally intense than are men's." Tyler (1956) hypothesizes a general attitude of wanting a career merely for the pursuit of any pleasant, congenial activity which offers itself until marriage, and McArthur (1958) suggests a "being"—rather than a "doing"—orientation. Warren (1959) summarizes other reasons: Women appear to have less clearly defined interests than do men, women are less intrinsically motivated, and women express their interests through more than one role. Harmon (1967b) concludes that women have a basic lack of vocational interest, but hastens to add that the pendulum is swinging away from work being described as a stopgap, toward work as an ongoing way of life (Harmon & Campbell, 1968).

In spite of the problems, the women's form of the SVIB continues to be used to clarify women's vocational interests. Perhaps the most efficient way to understand what interests are being measured is through factor analysis. Several factor analyses of the women's SVIB have been reported in the literature (Anderson, 1965; Crissy & Daniel, 1939; Darley, 1941; Strong, 1943). One of the common characteristics of these studies, however, is that they all made use of occupational scale scores as the variables to be factor analyzed. Such a procedure, among

¹ Completed in partial fulfillment of the requirements for the doctoral degree under the overall supervision of Edwin C. Lewis and statistical supervision of Leroy Wolins.

² Requests for reprints should be sent to the author who is now at the Counseling and Testing Center, Schofield Hall, University of New Hampshire, Durham, New Hampshire 03824.

others, has been strongly criticized by Guilford (1952) as being inappropriate.

The most probable reason for using scale scores, rather than the responses to individual items (as recommended by Guilford), was simply the lack of technical and computational facilities to handle such a large number of variables. Such self-limiting boundaries also restricted the one study located by the author that did utilize items rather than scales in factor analyzing the women's SVIB (Gernes, 1940).

Gernes' Ss were 500 Nebraska Teachers College freshman women. She selected 163 of the 410 items on the women's SVIB and clustered them according to the item-groupings on the inventory. Factor analysis of the 8 resulting groups yielded a total of 35 unlabeled factors. Although it would perhaps have been easier and even less time-consuming rationally to determine the subgroupings of 163 items clustered into 8 groups, the value of the Gernes study lies in the way in which she approached the problem, operating as effectively as possible within the limits of highly restrictive computational techniques. Her research, therefore, provided an impetus for the present study.

METHOD

The responses of 671 women-in-general to each of the 400 items on Form T400R of the SVIB were intercorrelated and the items grouped into 26 rationally coherent clusters, on the basis of the magnitude of their intercorrelations.³ The multiple group method of factor analysis (Thurstone, 1947) was followed, and 1 factor was extracted from each of the clusters. The angular cosines among the 26 oblique factors were then obtained and in turn factor analyzed, using a variation of Lawley's maximum likelihood procedure (Jöreskog, 1967). The resulting 9 common factors and the remaining 17 more specific factors were then rotated orthogonally. Finally, the factor loadings on the 26 factors of each of the 400 items were computed, and a 400 × 400 residual table was obtained.

RESULTS

Most of the 9 group factors (lettered A through I) were defined by items many of which also loaded on 1 or more of the 17

subgeneral factors (numbered 1-17). The following group factors were defined by items at least half of which also had substantial loadings on the indicated subgeneral factors: on A—6,7,10,11,13, and 15; on B—3; on C—6 and 10; on D—8 and 17; on E—3,4,5, and 14; on F—4,7,16, and 17; on G—3, 14, and 15; on H—none; on I—7.

Of the 400 items in the analysis, only 63 failed to load at least as high as $\pm .20$ on any of the 26 factors.⁴

Subgeneral Factor A

Of the 116 items with loadings of at least $\pm .20$, those with the highest positive loadings indicate a broad interest in designing costumes (.89) and children's clothes (.80), interior decorating (.65), fashion modeling (.53), illustrating (.52), managing a women's style shop (.52), decorating a room with flowers (.50), teaching art (.49), dressmaking (.47), displaying merchandise in a store (.47), being a beauty specialist (.46), buying merchandise (.46), being a florist (.46), and experimenting with new beauty preparations (.45). Middle-range positive loadings indicate an interest in such things as being an actress (.30), dramatist (.37), music composer (.29), opera singer (.29), professional dancer (.40), or radio-TV singer (.32), acting as a cheerleader (.30), and entertaining others (.31). These loadings, along with those in the lower range, seem to indicate a general aesthetic interest and would seem to be consistent with a summary label such as "Artistic Endeavors and Performing Arts."

Subgeneral Factor B

Sixty-five items have loadings of at least $\pm .20$. Interest in being a surgeon (.86), physician (.78), bacteriologist (.54), or biologist (.53), watching an open-heart operation (.53), being a nurse (.49), taping a bruised ankle (.48), and performing scientific experiments (.47), as well as being a member of any profession varying from criminal lawyer (.31) to professional golfer (.20),

³ The author wishes to express his appreciation to David P. Campbell of the Center for Interest Measurement Research, University of Minnesota, for making these data available.

⁴ Due to the pagination involved, only those items which were responsible for naming each factor will be listed and discussed. Factor loading tables for each of the 26 factors, as well as the factor loading matrix, are available from the author.

makes the label "Medical and Professional Endeavors" seem logical.

Subgeneral Factor C

Fifty-seven items help to describe this factor. The highest positive loadings indicate an interest in being a musician (.80), music composer (.74), or opera singer (.56), and playing the piano (.51). Lower positive loadings reveal an interest in art (.28, .21, .23, .29, .33, .28, .32, .39, .28), drama (.30, .39, .31, .29), poetry (.43, .35), and writing (.20, .24, .23, .30, .38, .20, .20). Clearly, the name "Music and Other Arts" would be appropriate.

Subgeneral Factor D

Sixty-three items are responsible for defining Factor D, and support interests in being a criminal lawyer (.70), corporation lawyer (.64), or judge (.63), being the governor of a state (.51) or a politician in general (.45), and being an employment manager (.32), life insurance saleswoman (.29), or hotel manager (.25). Therefore, the suggested label is "Law, Leadership, and Business."

Subgeneral Factor E

Factor E is defined by 80 items. Those with the highest positive loadings deal with being a statistician (.76) or bookkeeper (.75), studying statistics (.70), being an income tax accountant (.65) or computer operator (.61), making statistical charts (.61), and operating office machines (.58), among other quantitative, technical-type endeavors. On this basis, it is suggested that the factor be described as "Quantitative-Technical Clerical."

Subgeneral Factor F

Factor F is represented by 104 items. "White-Collar Work" seems to be the implied interest, inferred from such items as being a nurse's aid (.63), nurse (.50), dental assistant (.49), office clerk (.49), receptionist (.49), typist (.46), hospital records clerk (.45), and stenographer (.45).

Subgeneral Factor G

A scientific factor is indicated by the 43 items of Factor G. Such items as botany (.63), nature study (.55), zoology (.55), ge-

ology (.50), and bird watching (.46), as well as those with relatively low positive factor loadings, suggest that this factor would be most accurately called "Natural Science."

Subgeneral Factor H

Thirty-four items load at least $\pm .20$ on Factor H. Interest in physical education (.69), being an athletic director (.68), being a physical education director (.60), being a tennis champion (.53), and engaging in physical activity (.51) indicates that the factor should be named, simply, "Athletic."

Subgeneral Factor I

The 49 items comprising Factor I have, on the average, the lowest positive factor loadings of the nine subgeneral factors. The factor is sufficiently defined, however, to permit easy interpretation. Items pertaining to being an employment manager (.60), hotel manager (.48), or office manager (.46), plus other very similar items, are consistent with labeling the factor "Managerial."

Group Factor 1

Twenty-one items define Factor 1. Teaching kindergarten (.66), grade school (.55), and children in general (.51), and managing a children's nursery at a resort hotel (.45) strongly indicate interests in "Children and Teaching."

Group Factor 2

"Social Science" best describes Factor 2, summarizing the 9 relevant items; psychologist (.67), psychology (.61), and sociology (.35) are the items with the highest positive loadings.

Group Factor 3

Fifteen items define Factor 3. Being a scientific research worker (.57) or chemist (.50), performing scientific experiments (.49), and doing research work (.42) suggest the label "Research."

Group Factor 4

The 17 items which define Factor 4 are best summarized by "Secretarial": shorthand (.58), stenographer (.55), typist (.54), and private secretary (.50).

Group Factor 5

A "Quantitative" interest is indicated by the 6 items of Factor 5: algebra (.68), geometry (.55), arithmetic (.45), calculus (.37), physics (.28), and chemistry (.24).

Group Factor 6

Sixteen items are responsible for describing Factor 6; interest in being an artist (.58), liking prominent artists (.46), and teaching art (.45) suggest the title "Artistic."

Group Factor 7

Twenty-five items describe Factor 7. The items are bunched together in a relatively low range of factor loadings. Nevertheless, it seems most logical to interpret this factor as indicating interest in "Sales and Business." Retailer (.38), specialty saleswoman (.34), and waitress (.22) are representative items.

Group Factor 8

Factor 8 is represented by 14 items. Those dealing with politicians (.36, .63, .37, .27, .29, .30) and electioneering for office (.49), as well as several dealing with various outlets for speaking, define this factor as "Politics and Performing Arts."

Group Factor 9

Twelve items describe Factor 9 as having to do with civics (.69), political science (.57), and history (.47), in addition to other areas of the "Humanities."

Group Factor 10

Dramatist (.61), famous actress (.51), actress (.48), and dramatics (.47) are the most prominent of the 17 items that describe Factor 10 as a "Dramatics" factor.

Group Factor 11

Nineteen items load at least $\pm .20$ on Factor 11. The items that are most responsible for naming the factor "Literary" are author of best-selling novel (.59), magazine writer (.53), and author of novel (.50).

Group Factor 12

Items dealing with cooking (.49, .89), trying new recipes (.68), and preparing dinner

for guests (.33, .41, .56, .39) have the highest loadings of the 16 items that make it seem logical to interpret Factor 12 as being an indicator of "Culinary" interests.

Group Factor 13

All 4 items defining Factor 13 obviously represent a "Sewing" interest: preparing a meal versus making a dress (−.62), sewing (.56), dressmaker (.39), and home economics (.29).

Group Factor 14

The 14 items representing Factor 14 are concerned with repairing electrical wiring (.55), operating machinery (.51), inventing (.45), and tinkering (.45); they define the factor as "Mechanical."

Group Factor 15

An "Agrarian" interest is indicated for Factor 15 by those items, out of 10, with the highest loadings: planning the landscape (.48), landscape gardener (.43), and raising flowers and vegetables (.43).

Group Factor 16

Factor 16 clearly indicates a "Religious" interest, since practically all of the 15 items are directly related to spiritual matters: reading the Bible (.66), going to church (.62), religious people (.59), church worker (.57), and so forth.

Group Factor 17

Ten items are responsible for describing Factor 17. Since policewoman (.69) has the only relatively high loading, the factor would seem to warrant the name "Law Enforcement."

Residuals

Only three residuals had factor loadings greater than .35 (.38, .38, .43), and each comprised a couplet (correlated highly with only one other item). There were no residual factors.

Measurement Scales

"Homogeneous" factor-type scales were constructed by identifying those items which

can be considered to be relatively pure measures of each of the 26 factors. The items within each homogeneous scale have a high relationship to each other and a low relationship to the other items in the inventory.⁵

DISCUSSION

The vocational interests of women are indeed complex, not to be explained away by the reductionistic dichotomy of "career versus homemaking." In other words, the high specificity of structure—many identifiable factors—strongly suggests that women's vocational interests are far more complex than has been widely assumed.

The results of any analysis, however, are highly dependent upon the restrictions of the data to be analyzed. One restriction is that the dimensions of women's vocational interests are partially a function of the conception of those interests by those who construct the instrument to be used, in this case, the SVIB. For example, women were conceived to be interested in religion; several religion items were therefore included, which consequently yielded a "Religious" factor. Such was not the case with the men's form, however. (For comparison, the reader is referred to the Cranny [1967] factor analysis of the men's form which in many respects was a companion study to the present analysis.) In addition, the complexity of factors which resulted from the present analysis was based, to a large extent, on a sample composed of professional women.

The 26 factors isolated in the present study provide for a deeper understanding of the theoretical constructs underlying women's vocational interests, both in the complexity of those constructs and in the content of the specific factors. While a better understanding of the origin and development of women's interests has not been gained, it is meaningful that so many of the 26 factors are not directly vocationally oriented. Rather, many of the interests seem to be concerned with such things as homemaking and the feminine role. Perhaps this is a verification of the difficulty in measuring women's interests in general. Certainly, the present results reflect an

uncertainty of the role of interests in vocational choice, which is particularly true for women of college age, who do not tend to think in terms of long-term careers (Harmon, 1967a; Super, 1957).

Duntzman (1966) has commented on the communalities between two of the factor analyses (Anderson, 1965; Crissy & Daniel, 1939) which utilized scale scores rather than responses to individual items. The present study, however, stands alone and is not directly comparable with such studies. It is anticipated that the complex computer technology now available will be applied in similar ways in the future.

Research Implications

Clark (1960) has said that "homogeneous" measurement scales are the logical end result of a search for meaning such as the search made by the present study for more basic knowledge of the vocational interests of women. Having identified such scales, the question now arises as to their use. One application of the measurement scales would be to construct a short form of the women's form of the SVIB.

The possibility of the factor structure changing over time now needs to be recognized and should be investigated. This, along with an investigation of the predictive validity of the measurement scales and comparisons with replications of the present study and similar analyses of other interest measures, would seem to be possible as well as relevant. Overall, the results of the present study should facilitate future efforts toward the development of a sound theoretical structure of women's vocational interests.

REFERENCES

- ANDERSON, H. E., JR. A factorial study of the female form of the SVIB. *Journal of Applied Psychology*, 1965, **49**, 270-273.
- ARNS, J. A factorial analysis of the vocational interests of two hundred adult female students. Unpublished doctoral dissertation, Temple University, 1958.
- CLARK, K. E. Problems of method in interest management. In W. L. Layton (Ed.), *The Strong Vocational Interest Blank research and uses*. Minneapolis: University of Minnesota Press, 1960.
- CRANNY, C. J. Factor analytically derived scales for the Strong Vocational Interest Blank. Unpub-

⁵ Available upon request from the author.

- lished doctoral dissertation, Iowa State University, 1967.
- CRISSEY, W. J. E., & DANIEL, W. J. Vocational interest factors in women. *Journal of Applied Psychology*, 1939, 23, 488-494.
- DARLEY, J. G. *Clinical aspects and interpretation of the Strong Vocational Interest Blank*. New York: The Psychological Corporation, 1941.
- DARLEY, J. G., & HAGENAH, T. *Vocational interest measurement theory and practice*. Minneapolis: University of Minnesota Press, 1955.
- DUNTEMAN, G. H. A note on the factorial invariance of the female form of the SVIB. *Journal of Applied Psychology*, 1966, 50, 561-562.
- GERNES, E. A factorial analysis of selected items of the Strong Vocational Interest Blank for Women. Unpublished doctoral dissertation, University of Nebraska, 1940.
- GUILFORD, J. P. When not to factor analyze. *Psychological Bulletin*, 1952, 49, 26-37.
- HARMON, L. W. Women's interests—fact or fiction? *Personnel and Guidance Journal*, 1967, 45, 895-900. (a)
- HARMON, L. W. Women's working patterns related to their SVIB housewife and "own" occupational scores. *Journal of Counseling Psychology*, 1967, 14, 299-301. (b)
- HARMON, L. W., & CAMPBELL, D. P. Use of interest inventories with nonprofessional women: Stewardesses versus dental assistants. *Journal of Counseling Psychology*, 1968, 15, 17-22.
- HOGG, M. I. Occupational interests in women. *Personnel Journal*, 1928, 6, 331-337.
- JÖRESKOG, K. G. A general approach to confirmatory maximum likelihood factor analysis. Research Bulletin RB-67-48, Educational Testing Service, Princeton, New Jersey, 1967.
- KEYSERLING, M. D. Facing the facts about women's lives today. In *New approaches to counseling girls in the 1960's*. United States Government Printing Office, 1965.
- MCARTHUR, C. Comment. *Journal of Counseling Psychology*, 1958, 5, 48-49.
- MILLER, C. H. Vocational guidance in the perspective of cultural change. In H. Borow (Ed.), *Man in a world at work*. Boston: Houghton Mifflin, 1964.
- STRANG, R. Vocational interest blanks for women. In O. K. Buros (Ed.), *The 1940 mental measurements yearbook*. New Brunswick, N. J.: Rutgers University Press, 1941.
- STRONG, E. K., JR. *Vocational interests of men and women*. Stanford: Stanford University Press, 1943.
- SUPER, D. E. Strong's vocational interests of men and women. *Psychological Bulletin*, 1945, 42, 359-370.
- SUPER, D. E. *The psychology of careers*. New York: Harper, 1957.
- SUPER, D. E., & CRITES, J. O. *Appraising vocational fitness by means of psychological tests*. New York: Harper, 1962.
- THURSTONE, L. L. *Multiple factor analysis*. Chicago: University of Chicago Press, 1947.
- TYLER, L. E. *The psychology of human differences*. New York: Appleton-Century-Crofts, 1956.
- WARREN, P. A. Vocational interests and the occupational adjustment of college women. *Journal of Counseling Psychology*, 1959, 6, 140-147.

(Early publication received March 19, 1969)

OCCUPATIONAL GROUP AS A MODERATOR OF THE JOB SATISFACTION-JOB PERFORMANCE RELATIONSHIP¹

RICHARD E. DOLL² AND E. K. ERIC GUNDERSON

Navy Medical Neuropsychiatric Research Unit

San Diego, California

Personnel comprising wintering-over parties at small scientific stations in Antarctica represent two broad but quite different occupational groups: civilian scientists and Navy enlisted men. The motivations of the Navy enlisted men who volunteer are less related to their specific jobs in the Antarctic than are those of the civilian scientists. The results confirmed the hypothesis that occupational group is a moderator of the job satisfaction-job performance relationship, and that the relationship is more pronounced for the scientist group than for the Navy enlisted group.

One of the more unequivocal conclusions that has been drawn from past research on employee attitudes and job performance is that there is not a simple relationship between the two (Brayfield & Crockett, 1955; Kahn, 1960; Katzell, Barrett, & Parker, 1961) and that not much is likely to be learned from simple two-variable designs. Katzell et al. (1961) suggest that job satisfaction and performance should be considered as separate outputs of the work situation which, depending upon the intervening variables of employee needs and expectations, may or may not be correlated.

Differences in employee needs and expectations seemed to be reflected in occupational levels. Centers and Bugental (1966) demonstrated that, at higher occupational levels, intrinsic job components (opportunity for self-expression, interest-value work, etc.) were more valued while, at lower occupational levels, extrinsic job components (pay, security, co-workers, etc.) were more highly regarded.

Tannenbaum (1966) notes that one of the weaknesses in the hypothesis that associates productivity with satisfaction is the failure to distinguish between satisfaction and motivation. He points out that a person may be satisfied with his work insofar as his needs are met but his satisfaction indicates little about his motivation to work, particularly when his satisfaction does not depend on the amount of effort he puts into his work.

Personnel comprising wintering-over parties at small scientific stations in Antarctica represent two broad but quite different occupational groups. One group consists of civilian scientists whose sole *raison d'être* revolves around their individual scientific project. The other broad occupational group is composed of United States Navy enlisted men whose reasons for participating tend to be more diverse and less specific. Questionnaire data have shown that these men offer such reasons for volunteering as saving money, adventure, promotional opportunities, experience, etc. In other words, the motivations of the Navy enlisted men would seem to be less related to their specific jobs in the Antarctic (e.g., construction, mechanics, electronics) than those of civilian scientists.

The purpose of the present study was to investigate the hypothesis that occupational group, as defined herein, is a moderator of the job satisfaction-job performance relationship and that the relationship will be more pronounced for the scientist group than for the Navy enlisted group.

¹ Report Number 69-11, supported by the Bureau of Medicine and Surgery, Navy Department, under Research Work Unit MF12.524.003-9005. Opinions expressed are those of the authors and are not to be construed as necessarily reflecting the official view or endorsement of the Department of the Navy. The assistance of Mr. David H. Ryman is gratefully acknowledged.

² Requests for reprints should be sent to Richard E. Doll, Department of the Navy, Navy Medical Neuropsychiatric Research Unit, San Diego, California 95152.

METHOD

Subjects

The sample consisted of 129 Navy enlisted men and 66 civilian scientists who were assigned to United States Antarctic stations for one year. All Ss were volunteers who had been selected for Antarctic duty primarily on the basis of occupational competence. The Navy enlisted group was largely composed of Navy construction personnel ("Sea-bees") and Navy technical and administrative personnel.

The mean ages and educational levels for the two occupational groups were as follows: Navy enlisted—mean age 27.5 yr. and mean education 11.3 yr.; Scientists—mean age 25.7 yr. and education level at least a BA.

Procedure

The Ss filled out biographical, personality, and attitude inventories as part of the psychiatric assessment program for Antarctic volunteers. Along with specially constructed attitude inventories, performance measures consisting of supervisor ratings and peer nominations were obtained on two occasions during the year of duty in Antarctica, early in the period of winter isolation, and again near the end of winter isolation approximately 6 mo. later.

The data for the present study were based on two self-report scales reflecting job satisfaction and three criterion scores of job performance as derived from supervisor ratings, peer nominations, and a combination of the two. The job satisfaction scores and the job performance scores were obtained near the end of the winter isolation.

The job satisfaction scales were originally developed from a series of factor analyses of a set of attitude items administered to combined samples of Ss. These analyses produced two factors which seemed to fall under the rubric of job satisfaction. The finding of two factors is congruent with the conclusion of Hulin and Smith (1965) who point out that job satisfaction is not a unidimensional variable but is made up of a number of factors or distinct areas of satisfaction.

As differences between the Navy and civilian occupational groups became apparent, intercorrelation matrices of the scale items were examined for each group separately. The matrices were found to be very similar and only minor changes in scale content were required to make the job satisfaction scales comparable for the two groups.

The first scale consisted of five items reflecting a rather general expression of satisfaction or dissatisfaction with the wintering-over experience and was labeled Job Morale. This scale consisted of the following items: (a) I would like to go on another Antarctic expedition in the future. (b) I am happier with my job here than I was in my last assignment. (c) Time passes too slowly in the Antarctic (response value reversed). (d) I wish I could stay in the Antarctic longer than now planned. (e) I often wish I had never come to the Antarctic (response value reversed).

The second scale consisted of four items and was labeled Job Importance. The Job Importance Scale was composed of the following items: (a) My present duties employ my abilities in the best possible way. (b) The success or failure of this station depends as much on my job as any other. (c) My job here is important enough to justify my spending all this time in the Antarctic. (d) I feel that I am contributing as much on this expedition as others are.

A 6-point scale, anchored at the ends by "strongly disagree" and "strongly agree" was used for each item with a range from 1 to 6, with reversals where appropriate. The scale scores were the algebraic sums of the item weights.

The job performance criteria were based on independent ratings by two station leaders (supervisors) and nominations by all peers, both military and civilian. Each man assigned to the station was rated by the supervisors for the characteristics of "industriousness," "motivation," and "proficiency" according to an 8-point rating scale format. The peer nominations consisted of each man nominating the five most outstanding men on "industriousness." Criterion scores were converted to *T* scores (*M* = 50; *SD* = 10) without regard for occupational groupings. The third job performance criterion score was obtained by combining the supervisor and peer criterion scores.

Pearson correlations were computed between the two job satisfaction scores and the three criterion scores for the Navy enlisted group and the civilian scientist group separately and combined. Means and standard deviations for the various comparisons also were computed.

RESULTS AND DISCUSSION

The results presented in Table 1 confirm the hypothesis that measures of job satisfaction and job performance would be more closely related within the civilian scientist group than within the Navy enlisted group.

TABLE 1
CORRELATIONS BETWEEN JOB PERFORMANCE RATINGS AND JOB SATISFACTION SCALES

| Ratings | Navy (<i>N</i> = 129) | | Scientist (<i>N</i> = 66) | | Combined (<i>N</i> = 195) | |
|------------|---------------------------|----------|-------------------------------|----------|-------------------------------|----------|
| | <i>A</i> | <i>B</i> | <i>A</i> | <i>B</i> | <i>A</i> | <i>B</i> |
| Supervisor | .05 | .03 | .34** | .39** | .12 | .15* |
| Peer | .03 | .04 | .27* | .47** | -.09 | .14* |
| Combined | .04 | .01 | .33** | .44** | .03 | .16* |

Note.—*A* = Job Morale Scale; *B* = Job Importance Scale.
* *p* < .05.
** *p* < .01.

None of the job satisfaction–job performance correlations within the Navy enlisted group approached significance ($r \geq .14$; $p \leq .05$), while within the civilian scientist group all correlations were significant.

By not separating the two groups, one would conclude from the correlations shown under the combined group that there was little, if any, relationship between job performance and job morale, and a barely significant relationship between job importance and job performance. The ambiguity in interpretation resulting from combining two such widely different occupational groups is exemplified by the negative correlation obtained between the Job Morale Scale and the job performance peer ratings despite the fact that a positive correlation existed in both the Navy enlisted and scientist group comparisons. This paradox can be explained by examining the means of the peer ratings and the Job Morale Scale shown in Table 2; the Navy enlisted group had a higher peer rating score than the scientist group but a lower Job Morale score. When the two groups were combined, a negative relationship appeared. The standard deviations shown in Table 2 strongly suggest that the differences in correlations given in Table 1 are not simply restriction of range artifacts.

In addition to the above analysis, correlations were computed between the two job satisfaction scales for the two occupational groups separately. The correlation was .19 for the Navy enlisted group and .38 for the scientist group. This is a rather substantial difference and would indicate that occupational groups, as used in this study, might not only be a moderator of the job satisfaction–job performance relationships, but may also be a moderator of the job importance–overall job satisfaction relationship.

The results of the present study support the notion that the job satisfaction–job perform-

TABLE 2
MEANS AND STANDARD DEVIATIONS OF JOB PERFORMANCE SCORES AND JOB SATISFACTION SCALES

| Variable | Navy | | Scientist | | Combined | |
|------------|----------|-----------|-----------|-----------|----------|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Supervisor | 50.5 | 8.9 | 49.7 | 10.4 | 49.9 | 9.4 |
| Peer | 52.8 | 8.9 | 46.1 | 7.0 | 50.0 | 9.0 |
| Combined | 51.3 | 8.0 | 47.8 | 8.0 | 49.8 | 8.3 |
| Morale | 17.6 | 5.0 | 23.4 | 5.0 | 19.4 | 5.6 |
| Importance | 17.2 | 4.6 | 16.0 | 4.6 | 16.6 | 4.8 |

ance relationship is not a simple bivariate function. It was found that job performance ratings, both supervisor and peer, are substantially related to measure of job satisfaction for a group whose motives for being in Antarctica were primarily oriented toward specific tasks and goals but do not relate at all within a group whose motivations for volunteering are more diverse and less highly related to the job itself.

REFERENCES

- BRAYFIELD, H. H., & CROCKETT, W. H. Employee attitudes and employee performance. *Psychological Bulletin*, 1955, 52, 396–424.
- CENTERS, R., & BUGENTAL, D. E. Intrinsic and extrinsic job motivations among different segments of the working population. *Journal of Applied Psychology*, 1966, 50, 193–197.
- HULIN, C. L., & SMITH, P. C. A linear model of job satisfaction. *Journal of Applied Psychology*, 1965, 49, 209–216.
- KAHN, R. L. Productivity and job satisfaction. *Personnel Psychology*, 1960, 13, 275–287.
- KATZELL, R. A., BARRETT, R. S., & PARKER, T. C. Job satisfaction, job performance, and situational characteristics. *Journal of Applied Psychology*, 1961, 45, 65–72.
- TANNENBAUM, A. S. *Social Psychology of the Work Organization*. Belmont, California: Wadsworth, 1966.

(Early publication received April 14, 1969)

RECRUITING DOOR-TO-DOOR SALESMEN BY MAIL

VALENTINE APPEL¹

AND

M. R. FEINBERG²

Grudin/Appel Research Corporation

Baruch College, City University of New York

A research program involving over 20,000 door-to-door salesmen resulted in the development of an objectively scored mail questionnaire which effectively eliminated three-fourths of those who failed to meet minimum standards of sales performance. The content of the questionnaire and the scoring key clearly indicated that the personal characteristics which made for sales success were contrary to what the company had previously believed.

The company which is the subject of the article is a direct seller, that is, a company which distributes its merchandise by means of door-to-door salesmen. The National Association of Direct Selling Companies, the trade association for the industry, reports that there are about 1,200 such companies doing business in the United States; and although the total volume of business done by these companies is not precisely known, the total volume of the 200 members of the association is approximately five billion dollars annually.

Like many direct sellers, this particular company recruits its sales force almost entirely by mail. Every week about 2,000 teenage boys apply for jobs as salesmen in response to direct mail solicitation and to newspaper and magazine advertisements. Applicants are then mailed sales materials following examination of their application blanks and a delinquency check of the company records. For many years the application blank consisted only of the applicant's name and address, age, sex, and a parental endorsement; and virtually all of the applicants who applied were accepted. Although this company has long been established and operates a very profitable business, under this system a large majority of the boys who were started each week failed to perform to the company's minimum standard of sales performance. A sales failure was defined as someone who did not return a sales volume more than sufficient to recover the cost of recruiting and the mailing of materials. Generally this was a person who

voluntarily discontinued selling in the first month.

The situation presented an unusual research opportunity for improving the efficiency of the company's selection procedures. To add to the gravity of the problem, there are very few organizations—outside of the armed forces—where such large samples of personnel are continuously recruited to perform exactly the same job. Moreover, data were economically available. Since the company policy is to recruit through the mails, mail questionnaires were mandatory, and there was no question of nonrespondent bias, which is a major problem in most mail surveys.

THE RESEARCH PROGRAM

The research naturally divided itself into four basic parts: (a) a preliminary qualitative investigation, (b) the development of questionnaire items, (c) tests using nonreturn of the questionnaire as a screening device, and (d) revision and validation of the questionnaire items. Although each phase of the program has been duplicated one or more times as more data were required, for simplicity of presentation only one study from each of the four phases will be described here.

Qualitative Investigation

The initial step was to conduct a series of informal personal interviews in various parts of the country among boys who had been previously classified as successful or unsuccessful. In addition, prior to interviewing the salesboys, personal interviews were conducted with members of the company's sales department having responsibility for recruitment and selection. As a result of the investigation,

¹ Requests for reprints should be sent to Valentine Appel, Grudin/Appel Research Corporation, 105 Madison Avenue, New York, New York 10016.

² Also at BFS Psychological Associates, Inc.

it became obvious that most of the company executives believed that the potentially successful salesboy was one who came from a lower income home in which money was not easy to come by, and where it was necessary for the children to earn their own spending money. The following quotations, each from a different interview, will illustrate:

A boy whose family is . . . preferably in the lower income brackets. Generally speaking, boys from higher income families can get money much easier than by selling.

A boy from a home with a number of brothers and sisters . . . from a home where his parents cannot give him too much. If he wants money, he must find a way to get it himself.

A boy from a broken home who needs to find a way of life . . . a boy who wants additional money because of limited parental income.

In contrast to these Horatio Alger expectations, the interviewers who visited the boys in their homes reported that the successful boys appeared to be ordinary middle class teenagers. The unsuccessful boys, however, showed a marked contrast. They were difficult to locate, because they did not appear to live in one place for any length of time; and the homes in which they had previously lived appeared typically to be run down. When the children were located, they were frequently found in households which had been deserted by one parent or the other, and in general their families were depressed economically.

Questionnaire Development

On the basis of the preliminary qualitative investigation, a series of eight multiple-choice questions was prepared for use in a mail questionnaire. These questions were specifically written to resolve the severe contradiction which existed between the prevailing opinion in the sales department and the reports of the interviewers who contacted the salesboys in the field.

Because of the sensitive nature of the data which were required, plus the fact that a mail questionnaire was to be employed, some disguise in questioning appeared to be in order. Accordingly, questions were written which were known or believed to be correlated with family cohesiveness and socioeconomic status, but which on the surface appeared not di-

TABLE 1
SOME OF THE DISCRIMINATING QUESTIONNAIRE
RESPONSES

| Questionnaire responses | Successful applicants | Unsuccessful applicants |
|--------------------------------------|-----------------------|-------------------------|
| Has a bicycle | 75% | 52% |
| Family receives a newspaper | 72% | 52% |
| Attended show or circus with parents | 64% | 43% |
| Has a telephone at home | 63% | 38% |
| Signed application in pen | 59% | 46% |
| Saves money earned | 57% | 38% |
| Family less than 6 children | 54% | 30% |
| Base: all returns ^a | 195 | 365 |

^a Actually these data are from a later mailing—a sample independent of the one upon which the original analysis was conducted and which confirmed the original findings.

rectly related. Such questions related to bicycle ownership, telephone subscription, participation with parents in recreational activities, use of money earned, and size of family.

The questionnaires were mailed to all the boys (approximately 2,000) who had been started as salesboys in a given week. By the close of tabulation exactly two-thirds of the questionnaires had been returned. The sample was then divided into random halves on an every other questionnaire basis. Using one of the random halves, the questionnaire responses of the successful boys were compared with those who were less successful.

This comparison, which is illustrated in Table 1, essentially confirmed the findings of the original qualitative study in which the boys were visited in their homes. Moreover, the data contradicted the prior opinions of the sales executives who believed that sales success came of economic need. Paradoxically, the reverse appeared to be true. Those who appeared least in need of money also appeared most likely to exert themselves to earn it. The successful boy, when compared with his unsuccessful counterpart, was more likely to own a bicycle (75% vs. 52%), participate in recreational activities with his parents, have a telephone at home, come from a family of less than six children, and so forth.

On the basis of this comparison, a unit-weighted scoring key was developed in which those responses which were characteristic of successful boys were scored +1 and those

TABLE 2

RELATIONSHIP BETWEEN QUESTIONNAIRE
RESPONSE AND JOB SUCCESS

| Applicants | Failed to return | Returned questionnaire and scored in | | |
|----------------------|---------------------|---|-----------------|--------------|
| | | Bottom third | Middle third | Top third |
| Successful | 8% | 22% | 46% | 59% |
| Unsuccessful | 92% | 78% | 54% | 41% |
| Base: all applicants | 309 | 221 | 205 | 198 |

which were characteristic of unsuccessful boys were scored -1. All other responses were scored zero. This key was then applied to the remaining half-sample which had not been examined to this point. The total scores on the questionnaire were then related to whether or not each boy met the company's minimum standard of sales success.

The results of this analysis are summarized in Table 2 which divides the second half-sample into four groups: those who failed to return the questionnaire, and those who did return and who scored in the top, middle, and bottom thirds using the scoring key which had been independently developed using the first half-sample. By this device the percentage of successful applicants was increased from 8% to 59%, depending upon whether or not the questionnaire was returned and how well it scored.

Nonreturn as a Screening Device

Our intention in developing the questionnaire items was to produce material for inclusion in the application forms which were to be part of the recruitment ads. It was never the intention to send a questionnaire to sales applicants as a second and separate mailing. The possibility, therefore, of using a second mailing and of screening out all those who failed to return the questionnaire, as implied by Table 2, was frankly unexpected. Moreover, owing to the high cost of recruitment plus the limited availability of applicants, there was resistance by the sales department to the use of a second mailing because of the possibility that a large percentage of successful boys might be lost through their failure to return the questionnaire. In addition, there is a common belief among direct sellers that even a brief delay in placing sales material

in the hands of sales applicants reduces the probability of their being successful. Any delay is presumed to reduce the motivation of the applicant.

There was the additional question of whether the failure of the unsuccessful applicants to return the questionnaire would be predictive of their later failure or whether the nonreturn was caused by their prior failure. After all, the questionnaires were received by the applicants following receipt of merchandise to sell, and it was entirely likely that the decision not to return the questionnaire was prompted by their prior sales failure.

In order to resolve these doubts a special test was conducted involving 483 sales applicants who were divided into two matched samples on an every-other-name basis. One sample—the screened sample—was mailed a questionnaire and then merchandise to sell only upon return of the completed questionnaire. The unscreened sample was mailed merchandise immediately as per the usual procedure, and no questionnaire was sent at all. The results of this test are summarized in Table 3 from which two conclusions may be drawn: (a) The screening did not reduce the percentage of successful applicants (10% in the screened sample and 9% in the unscreened sample), and (b) the use of the questionnaire as a screening device had effectively eliminated half the unsuccessful applicants (45% out of 90%).

Armed with this information, the company adopted the technique of using a separate mail questionnaire. The nonreturn of this questionnaire was then employed as a screening device.

TABLE 3

NONRETURN OF THE QUESTIONNAIRE
AS A SCREENING DEVICE

| Type of applicant | Screened sample | Unscreened sample |
|---------------------------|--------------------|----------------------|
| Successful | 10% | 9% |
| Unsuccessful | 45% | 91% |
| Screened out ^a | 45% | — |
| Base: all applicants | 241 | 242 |

^a Questionnaire not returned.

Questionnaire Revision

To this point the scoring of the questionnaire was not included as a step in the selection procedure for the same reason that the company was initially reluctant to use the nonreturn of the questionnaire as a screening device. The limited supply of applicants and the need to maintain sales volume made it necessary to maintain a minimum sales staff, however inefficient that sales staff might be. Moreover, because the scoring key had been developed upon a sample of boys who may have already experienced some measure of success or failure prior to having received the questionnaire, there was a natural reluctance to adopt the scoring key under conditions where the questionnaire had to be returned before the merchandise could be sent.

However, since use of the new questionnaire had now been adopted as standard procedure, it became a simple matter to conduct additional screening studies to add questionnaire items, to delete others, and to improve the scoring key. Finally, as a result of a series of tests involving over 20,000 boys, a 7-item questionnaire, printed on a $4\frac{1}{2} \times 8$ business reply card, was developed to be mailed out on receipt of a sales application. On the basis of these tests the questionnaire was demonstrated to predict with sufficient accuracy as to be made operational which applicants would later be successful and which would not.

The data from the most recent such tests are summarized in Table 4 which divides 560 applicants who returned the questionnaire into two groups: those who were successful and those who were not. For each of these two groups the percentages of applicants scoring in top, middle, and bottom thirds is shown. From this, it can be seen that by rejecting all those who failed to score in at least the middle third, 44% of the potentially unsuccessful candidates who returned the questionnaire at the cost of 16% of the success could be eliminated.

The estimated utility of the entire program, including the use of nonreturn of the questionnaire as a screening device plus the additional screening following scoring of the questionnaire, can be calculated from the

TABLE 4

PERCENTAGES OF SUCCESSFUL AND UNSUCCESSFUL APPLICANTS REJECTED AT THREE CUTTING POINTS

| Cutting point | Successful applicants | Unsuccessful applicants |
|--------------------------------|-----------------------|-------------------------|
| Top third | 57% | 27% |
| Middle third | 27% | 29% |
| Bottom third | 16% | 44% |
| Base: all returns ^a | 195 | 365 |

^a Individual item responses reported in Table 1.

data presented in Tables 3 and 4. Table 3 determined that nonreturn of the questionnaire eliminated 50% of the unsuccessful applicants at a cost of virtually none of the successes. Table 4 further determined that if all questionnaire returnees who scored in the bottom third were rejected, 44% of the failures who were not screened out on the basis of questionnaire nonreturn would be eliminated. Since we estimate that 50% of the potentially unsuccessful applicants were already screened out by their failure to return the questionnaire, it follows that 44% of unsuccessful returnees represents an additional 22% of all unsuccessful applicants. From this comes the estimate that the program was able effectively to eliminate about three-fourths (50% plus 22%) of the sales failures at a cost of about one-fifth (16%) of the successes.

DISCUSSION

Although the use of biographical data in the selection of sales personnel has a history dating back nearly 50 yr. (Goldsmith, 1922; Harrell, 1960; Kornhauser, 1941; Kurtz, 1941; Manson, 1925), the present study provides an unusually clear example of how the use of an objectively scored questionnaire, economically administered through the mail, was able effectively to reduce sales and marketing costs. The study again points up the substantial discrepancy which sometimes exists between management opinion and what the market facts actually are. A similar discrepancy was previously demonstrated in a study by Blum and Appel (1961) relating to package design. That study showed the com-

plete inability of industrial designers and marketing management to predict consumer tastes and perceptions. The present study demonstrates how seriously the discrepancy between what the facts actually were and what they were believed to be had seriously limited the effectiveness of the company's sales and marketing program.

REFERENCES

- BLUM, M. L., & APPEL, V. Consumer versus management reaction in new package development. *Journal of Applied Psychology*, 1961, **45**, 222-224.
- GOLDSMITH, D. B. The use of the personal history blank as a salesmanship test. *Journal of Applied Psychology*, 1922, **6**, 149-154.
- HARRELL, T. The validity of biographical data items for food company salesmen. *Journal of Applied Psychology*, 1960, **44**, 31-33.
- KORNHAUSER, A., & McMURRY, R. N. A scientific prediction for the selection of salesmen. *Journal of Applied Psychology*, 1941, **24**, 11-17.
- KURTZ, A. Recent research in selection of life insurance salesmen. *Journal of Applied Psychology*, 1941, **25**, 11-17.
- MANSON, G. E. What the application blank can tell. *Journal of Personnel Research*, 1925, **4**, 1-28.

(Early publication received July 1, 1969)

USE OF THE WORK COMPONENTS STUDY WITH NEW COLLEGE-LEVEL EMPLOYEES

ROBERT N. FORD

American Telephone and Telegraph Company

EDGAR F. BORGATTA

University of Wisconsin

AND GEORGE W. BOHRNSTEDT¹

University of Minnesota

Data based on new college-level employees are reported which show the reliability and structure of the Work Components Study (WCS). A comparison of those persons who remain with the company with those who leave at the company's initiative and with those who leave of their own initiative shows few differences in their orientations toward work as measured by the WCS. However, when the company's rating of how many years it will take the new hires to reach the third level of management was taken as a criterion, it was found that those hires who score highest on WCS Score 3, Competitiveness desirability, highest on the SCAT verbal ability score, and highest on a personality measure of responsibility are those perceived by the company as moving ahead fastest. These results suggest that it is not the "organization man" type who is likely to be promoted, but the man perceived to be highly competitive, intelligent, and responsible.

In a recent paper, a revised version of the Work Components Study (WCS) was described (Borgatta, Ford, & Bohrnstedt, 1968). Data were reported on the between- and within-cluster median intercorrelations for the scores, as well as reliability statistics for independent replications with large samples of male and female freshman college students. Additionally, information was reported on the relationship of the WCS to personality inventory scores, college entrance test scores, and educational and income aspirations. In the current paper, use of the WCS with large groups of college-level personnel hired by a major industrial organization is reported. Data are presented on the reliability of WCS scores with these samples. However, the central concern of this paper is the exploration of the predictive power of the WCS to employment outcomes for a subsample of new college-level employees.

On a voluntary basis, new college-level employees of companies in the Bell Telephone System completed a questionnaire that included the WCS. The employees were given the questionnaire before actual arrival on the job (e.g., plant visit) or as soon after arrival

as was convenient. Cooperation was encouraged, but the emphasis on confidentiality and the voluntary basis of cooperation was stressed. Questionnaires were not anonymous, but the instructions noted that the individual data would in no way be made available to the cooperating companies and could not affect the career patterns of the respondent. Questionnaires were accompanied by addressed envelopes to be mailed directly to the researchers at the University of Wisconsin. It is estimated that a substantial majority of the new employees sent in questionnaires, but it was not possible to determine the actual proportion of cooperating employees because of differences in level of cooperation of the companies themselves.

Usable data were available for 869 male and 344 female college-level personnel hired during 1964. Employees separated from the company were classified by whether the company initiated the separation or the employee initiated it. Further, there were notations in the company's record as to whether the progress of the employee was satisfactory or unsatisfactory at the time of separation, but this distinction will not be maintained here since, in general, employees separated at the initiative of the company must have been less than satisfactory in some respect. Thus, the three

¹ Requests for reprints should be sent to Edgar F. Borgatta, Social Science Building, University of Wisconsin, Madison, Wisconsin 53706.

groups of employees considered were those still with the company after at least 1 yr., those separated by the company (or those rated unsatisfactory for the company at time of separation), and those who left the company at their own initiative but whose progress was rated as satisfactory at the time of separation.

A Description of the WCS

The theoretical basis of the *WCS* builds particularly on the work of Herzberg, Mausner, and Snyderman (1959), which posits a two-level theory of motivation. The first-level factor is based on the objective elements of the job situation itself, (for example, recognition, possibility for growth, working conditions, job security, etc.), which result in good or bad feelings about the job. The second-level factors refer to the individual's attempt to relate the job situation to his own felt needs (e.g., need for recognition, need for growth, etc.). In discussing feelings of unhappiness, the authors suggest that it is not the job itself, but the conditions which surround the job which are important in determining satisfaction.

In developing the *WCS*, attention was paid to Herzberg's factors, especially those which indicated concern with conditions surrounding the job. The actual development of items can be found in Borgatta (1967) and Borgatta et al. (1968). The revised version of the *WCS* contains seven scores. A brief description of each follows:

1. Potential for personal challenge and development (8 items). This score contains items which measure the desire to be in situations where there is an opportunity for creative work, a chance for as much responsibility as one wants, and where there is an emphasis on originality and individual ability.

2. Responsiveness to new demands (7 items). The items in this score determine the individuals' responsiveness to emergency situations in the job, changing job assessments, and such irregular demands.

3. Competitiveness desirability (and reward of success) (9 items). Here emphasis is on whether or not the individual seeks job situations where the salary is determined by

merit, competition is keen, and where there is emphasis on accomplishment.

4. Tolerance for work pressure (7 items). This score taps attitudes toward situations where the work load might be excessive, where a person might be on call after hours or might even have to take work home.

5. Conservative security (12 items). These items were designed to determine whether the individual wants to play it safe and have security. Items ask the importance of seniority, well-defined promotion guidelines, and well-defined job routines.

6. Willingness to seek reward in spite of uncertainty versus avoidance of uncertainty (12 items). Is the person willing to do interesting work even though he might get fired easily? Would he work with a company with interesting work even though it might be a short-run job? These are the kinds of attitudes explored in this score.

7. Surround concern (9 items). This score measures the respondent's concern with such "hygienic" aspects of the job as whether the lighting and ventilation are good, whether co-workers and supervisors are nice people, and whether the community has adequate cultural, social, and recreational opportunities.

Some Characteristics of the WCS

The data from all respondents were pooled and reliability estimates calculated. The estimates used were alphas (Cronbach, 1951) and are shown in the diagonal of the matrix in Table 1. It will be noted that the lowest coefficient is .66, and the coefficients range to .83, magnitudes indicative of reasonable reliability.

The intercorrelations for the *WCS* scores are also indicated in Table 1. For these samples of college hires, the intercorrelations among Scores 1, 2, 3, and 4 are somewhat higher than in the previous study with college freshmen (Borgatta et al., 1968), especially among the females. While these correlations are higher than are desirable, an equally important question is whether the individual scores differentially predict some criterion. As will be seen later, they do.

While the *WCS* scores are designed to measure various facets of work motivation, they might be defined as personality concepts if personality were broadly defined.

TABLE 1
RELIABILITIES AND INTERCORRELATIONS OF WCS SCORES

| Characteristic | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|--|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 1. Potential for personal challenge and development | <u>.67</u> .77 | .39 | .41 | .24 | -.10 | .13 | .21 |
| 2. Responsiveness to new demands | .54 | <u>.66</u> .67 | .52 | .62 | -.17 | .32 | .07 |
| 3. Competitiveness desirability (and reward of success) | .50 | .49 | <u>.70</u> .71 | .54 | -.24 | .43 | -.01 |
| 4. Tolerance for work pressure | .35 | .55 | .52 | <u>.79</u> .74 | -.15 | .35 | -.06 |
| 5. Conservative security | -.21 | -.23 | -.11 | -.13 | <u>.74</u> .78 | -.30 | .27 |
| 6. Willingness to seek reward in spite of uncertainty vs. avoidance of uncertainty | .13 | .23 | .28 | .30 | -.18 | <u>.81</u> .80 | -.19 |
| 7. Surround concern | .17 | .05 | .04 | .01 | .20 | -.18 | <u>.83</u> .83 |

Note.—Data above the main diagonal are for 869 male college hires; below the main diagonal for 344 female college hires. Cronbachs alphas are in the main diagonal.

However, it was demonstrated in Borgatta et al. (1968) that there is substantial content in the *WCS* which is independent of that contained in three personality inventories among a sample of college students. The three personality inventories—the Behavioral Self-Ratings (BSR; Borgatta, 1964), the Self-Identification Form (S-Ident; Borgatta, 1965), and the Interpersonal Orientations Form (IO; Borgatta & Bohrnstedt, 1968)—are short forms designed to tap personality content which is contained in longer forms. The titles of the scores are shown in Table 3. Based on data in the current report, the reliabilities of the S-Ident ranged from .53 to .81, the BSR from .63 to .91, and the IO from .55 to .85, again reasonable values.

The correlations between the *WCS* and the scores in the three personality forms in the current study replicated most of those obtained in the earlier study with college students indicating that there is substantial independent content in the *WCS*. The correlation matrix is excluded from the current paper, but copies may be obtained from the senior author. However, it was deemed important also to determine whether or not the personality measures and the *WCS* make independent contributions to the prediction of the evaluations made of the hires by the com-

pany. Additionally, it needs to be shown that both kinds of measures (work orientations and personality) can account for variation in ratings above and beyond that accounted for by measured ability. This analysis shall be presented below, but first the authors shall examine the relationship of the *WCS* to another kind of criterion, employment status of the hire after 1 yr. with the company.

WCS Scores and Employment Status at Least One Year After Hiring

The company's evaluations of new college-level employees were made approximately 1 yr. after joining the company. Information was obtained through a central service rather than directly through companies; thus the information was really available only on the basis of the "census" reports gathered for all companies. Since many factors are involved in reporting, data that were up-to-date were available only in part at the first relevant census, and additional information for the participants in the study was available only an additional year later. For the current comparison, the final information available is reported in Table 2, which presents mean scores for those still with the company, those separated from the company at the initiative of the company or who were noted as having

TABLE 2

WCS MEAN SCORES AND EMPLOYMENT STATUS AT LEAST ONE YEAR AFTER HIRING

| Characteristic | Still with company (<i>n</i> = 334) | | Separated from company | | | |
|--|---|-----------|--|-----------|---|-----------|
| | | | Company initiated or progress un- satisfactory (<i>n</i> = 56) | | Progress satisfactory (<i>n</i> = 30) | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| 1. Potential for personal challenge and development | 26.1 | 2.6 | 26.0 | 2.8 | 27.1 | 3.1 |
| 2. Responsiveness to new demands | 20.2 | 2.4 | 19.9 | 2.7 | 20.5 | 2.3 |
| 3. Competitiveness desirability (and reward of success) | 25.3 | 3.2 | 24.2 | 3.5 | 25.5 | 3.0 |
| 4. Tolerance for work pressure | 17.3 | 2.8 | 17.4 | 2.9 | 16.9 | 3.6 |
| 5. Conservative security | 15.4 | 4.7 | 15.4 | 4.7 | 14.0 | 4.9 |
| 6. Willingness to seek reward in spite of uncertainty vs. avoidance of uncertainty | 20.7 | 5.4 | 19.9 | 5.8 | 21.8 | 6.4 |
| 7. Surround concern | 24.2 | 3.2 | 24.5 | 3.4 | 24.2 | 3.8 |

Note.—There were 19 persons of indeterminate status (leave on armed services assignment, etc.) not in the above data.

initiated their own separation but who had a company rating of unsatisfactory progress at the time of separation (6 cases) and, finally, additional information is available for 30 employees who initiated their own separation and who had a company rating of satisfactory progress at the time of separation. Examination of the profile of scores for those still with the company and those separated from the company with progress unsatisfactory indicates that the differences in the profiles are not substantial. The only difference that occurs that would satisfy a statistical significance test (symmetric test at .05 level) is between those still with the company and those who left the company at the company's request on *WCS* Score 3, Competitiveness desirability, with the latter group scoring lower.

The group of 30 persons who left the company at their own initiative may be compared with those still with the company. Two differences in the profile may be noted. Those who left the company and were rated as having satisfactory progress were somewhat higher on *WCS* Score 1, Potential for personal challenge and development, and somewhat lower on *WCS* Score 5, Conservative security. However, the sample size involved is small, and

the results may be unstable. Indeed, neither of these differences would satisfy a symmetric test of statistical significance at the .05 level. Thus, at best, whatever difference occurs here may merely add fuel to the often self-critical attitudes of a company that it is losing some good men, but the facts cannot be confirmed in these data. Additional discussion of such projected differences will be found below.

SCAT Scores and Employment Status

It is instructive to examine the abilities scores of three groups as measured by the School and College Abilities Test (*SCAT*; Educational Testing Service, 1964). For the 334 cases still remaining with the company, the average *SCAT* score was 97.9. For the 56 who were separated from the company at the initiative of the company, or whose records indicated unsatisfactory progress, the average *SCAT* score was 95.2. This difference would satisfy a symmetric test of statistical significance at the .05 level. For the 30 persons who left the company at their own initiative and whose records indicated satisfactory progress, the average *SCAT* score was 98.1. Again, while these differences are not large, they are consistent with the expectations about the

selective processes once one is employed. Persons of lesser ability who are either rated unsatisfactory or who are separated at the initiative of the company tend to have lower SCAT scores. Employees who separate at their own initiative and who are displaying satisfactory progress at the time of termination appear to have characteristics similar to those who remain with the company.

It should be noted that the selection process with SCAT is already rigorous; that is, in general, persons who are hired have SCAT raw scores of 91 or higher; they are a select group from the point of view of measured abilities. Thus, the fact that even in so select a group persons of higher measured abilities still tend to succeed more is impressive.

Criterion Ratings as Related to WCS, SCAT, and Personality Scores

In this section two sample bases for analysis are utilized. One of the determinants of sample size for the current data was the availability of information concerning employment status or availability of Initial Management Development Program (IMDP) ratings, which are to be made after 1 yr. of service with the company. IMDP ratings are generally defined as the number of years estimated to be required for the employee to reach the third level of management. The first sample base is constituted of the 334 cases for which IMDP ratings were available in the second census following the year of hiring (hereafter called "IMDP after 2 yr." sample). It is not known what factors are associated with late reporting of IMDP ratings, but in order to get more complete information, delay of the analysis until the second census was required. Since late ratings were probably made later in the careers of Ss, at least one source of error arises since IMDP ratings are defined in terms of time. Thus, if there is a delay of a year in the IMDP rating, all persons are not being rated on exactly the same basis.

The second sample base is of 390 cases and includes those persons with IMDP ratings at the time of the second census and those persons positively identified as separated at the initiative of the company or at their own initiative under circumstances where the

company files indicate unsatisfactory progress (hereafter called "Inclusive" sample). In this case, the IMDP rating accorded to the employee is a value of 8, which is also the value coded for persons who were given a rating that they would never achieve the third level of management. Other than this, the coded values ranged from 2 to 7, and it should be known that the implicit expectation for those in IMDP is that the average person should be ready to carry a position at the third level of management in 5 yr. Thus, ratings tend to be relatively narrow in range as well as subject to the possible errors in terms of actual time when the ratings are made.

The IMDP ratings are, from the point of view of a company, an early indication of progress and potential. Further, in the sense that favorable impression leads to the implementation of favorable action—the self-fulfilling prophecy—the IMDP ratings should correspond to performance. Obviously, examination of what actually happens will require a subsequent follow-up, but in the interim the IMDP ratings may be revealing since they indicate what kinds of persons are perceived to be good by the company.

The IMDP ratings have a number of intrinsic limitations as criteria, however, which need to be elaborated. The restricted range of the ratings has already been noted, and, indeed, roughly 70% of ratings made indicate that the person will reach the third level of management in 4, 5, or 6 yr. Additionally, since the employees in this study are drawn from many companies, it has to be recognized that there are company differences, and these are relatively systematic biases. If in a company it is known that it takes somewhat longer to reach the third level of management than in another company, this will systematically influence the ratings. Further, independently of merit which may be rewarded in other ways, persons in some occupational categories may take longer to reach the third level of management and this may be due to overhiring tendencies or to a lesser frequency of vacancies than was built into the hiring formula. Thus, realistic differences in the way the reward system operates within the company may interfere with the effectiveness of IMDP ratings as a criterion.

A note of exact source of information is required with regard to the IMDP ratings. At the first census, IMDP ratings were available for very few hires. At the census of 1965, information was available for 286 persons. Thus, for 48 persons, the IMDP ratings became available after that census for various reasons. Indeed, the status of 19 persons appropriate for the study was still indeterminate at the time of this analysis for various reasons, such as leave of absence for the armed services, training, and so forth. The 334 cases with IMDP ratings in this part of the analysis, thus, are representative of all persons for whom a rating was available and who were still employed when the final check on data availability was made. Although the data are not presented here, the subsample of 286 cases for whom the IMDP ratings were available at the earlier point in time tend to show slightly higher relationships with the IMDP ratings, and the IMDP ratings have a slightly smaller standard deviation. Thus, even though the range of the ratings may be a little smaller at that earlier point in time, they appeared to bear slightly more relationship to the variables which are being used to predict the IMDP ratings. This might mean that the earlier set of ratings represents a more homogeneous set of standards and a more reliable criterion. However, these small differences will be ignored and the larger sample utilized.

IMDP ratings are less favorable the larger the value (since the value is the number of years estimated before readiness to perform well at the third level of management). Thus, in our analysis things that are negatively correlated to IMDP ratings are associated with the desirable side of the continuum.

Table 3 presents the correlations of IMDP ratings with *WCS*, *SCAT*, and personality scores which are derived from the forms completed voluntarily by the participants in this research.

Table 3 contains both zero-order and multiple correlations as well as the regression equations for estimating IMDP ratings from the *WCS* scores, *IO* scores, and personality measures. The table is broken into two parts (Columns 1-4 and 5-8) to coincide with the definitions of the IMDP after 2 yr. and Inclusive sample described earlier. Columns 1 and

5 contain the zero-order correlations between the 25 independent variables and IMDP ratings for the two samples. Columns 2 and 6 contain the standardized regression coefficients (SRC) for estimating IMDP scores from the *WCS* alone. Columns 3 and 7 show the SRCs for the regression of IMDP ratings on the *WCS* and *SCAT* scores. Finally, Columns 4 and 8 indicate the SRCs for estimating IMDP ratings from all scores (i.e., *WCS*, *SCAT*, and the three personality measures). The multiple correlations associated with each column are at the bottom of Table 3.

In this discussion the data for the IMDP after 2 yr. sample will be used as the point of departure. The data for the larger sample which includes also those no longer with the company who left earlier at the initiation of the company or with a notation on their records of unsatisfactory progress ($n = 390$) tend to be parallel to those of the smaller sample. It is left to the reader to examine the exact differences. In general, slight differences exist with a little less of the variance in IMDP ratings explained in the larger sample.

The largest single zero-order correlation coefficient with the IMDP ratings is *WCS* Score 3, Competitiveness desirability. Apparently, persons who indicate at the time they are employed that they want to get into situations where they have a chance of being rewarded for success and where they have an opportunity to show their abilities competitively are those who are receiving IMDP ratings of a fewer number of years, that is, are predicted to reach the third level of management earliest.

The second largest correlate with the IMDP ratings is Behavioral Self-Rating Responsibility (BSR), the content of which is orientation to task completion, responsibility, and conscientiousness.

Since there are a fair number of zero-order relationships beyond those discussed, examination of these will be left to the reader. The authors now proceed to the regression analysis examining the standardized regression coefficients (SRC) instead of the zero-order correlations.

Using the seven *WCS* scores only (Column 2), the multiple R is .31. This means that a large amount of variance is left unexplained,

but the circumstances under which prediction is occurring should be recalled as highly restrictive. For example, the sample is restricted in intelligence and motivation when hired. Further, the sample represents those who are succeeding after at least 1 yr. Under these circumstances, it can reasonably be asserted that any prediction is interesting and important. As Cronbach (1960, pp. 348-351) indicates, assessments which have relatively low correlations can be very beneficial to a company in its hiring if the selection ratio is not high (i.e., if the company has an ample supply of applicants), and if individual differences in the ability to perform the job are large. Cronbach notes that "coefficients as low as .30 are of definite practical value [p. 349]."

There are two statistically significant regression coefficients. First, *WCS* Score 3, Competitiveness desirability, has a standardized regression coefficient of .12. Thus, Conservative security is involved in a negative way in the assessment of performance, although the involvement is quite small.

Although not shown in Table 3, the importance of the motivational information may be judged in contrast to performance on SCAT scores. For the two subscores of SCAT (which are a more efficient predictor than the total score), the multiple R is .16, and virtually all of the variance is accounted for by the verbal score. Again, here it has to be emphasized that the group is explicitly select from the point of view of measured ability, not only in the general use of a SCAT score of 91 or higher in hiring, but also because those who have dropped out of the sample to this point have statistically significant lower SCAT scores. The persistent relationship of SCAT to the IMDP criterion is of particular interest from the point of view of validity of selection procedures. Since the motivational criteria in selection are less formally applied, essentially there is more opportunity for a relationship to be observed.

From the multiple R resulting from the regression of IMDP from both the seven *WCS* scores and the two SCAT scores, it is demonstrable that the predicted variance from the two types of information is relatively independent since the standardized regression co-

efficients remain almost identical in the combined computation. Further, the coefficient of determination of the overall multiple R is only .004 less than the sum of the two separate coefficients of determination.

It should be noted that the difference between $R = .16$ (the R based only on the SCAT scores) and $R = .31$ is significant at the .05 level by an F test indicating that the *WCS* does add predictive variance in IMDP ratings above and beyond that explained by SCAT scores above.

In order to facilitate interpretation of the relationship of the *WCS*, SCAT, and personality scores with regard to the IMDP ratings, a multiple R was computed, involving all these scores. These standardized regression coefficients are also to be found in Table 3 (Columns 4 and 8). The standardized regression coefficients that would be judged statistically significant, if one were testing hypotheses using a symmetric test with $\alpha = .05$, are italicized. Since a large number of variables are involved in the prediction, questions about the uniqueness of the sample are important in any interpretation. One needs to be relatively cautious, in other words, and some of the findings should be interpreted as of speculative interest rather than as hard facts. Still, it is of interest to see what is happening. Some of the relationships that might have been statistically significant as first-order r 's disappear. Other zero-order relationships which were not visible (suppressed) become visible. For example, the S-Ident intellectual orientation appears to increase in magnitude and become significant in the prediction equation. That is, concern with intellectual matters (in the sense of interests or activities that are normally classed as intellectual in nature) appears to be negatively associated with a favorable IMDP rating. The BSR responsibility score is positively associated with a favorable IMDP rating, and it should be noted that the SCAT verbal score appears to have greater prominence in the multiple R involving more variables. Apparently, then, some personality characteristics are acting as suppressors masking the relationship of SCAT and IMDP, which is stronger when the personality characteristics are taken into account.

TABLE 3
MULTIPLE CORRELATION OF WCS, SCAT, AND PERSONALITY SCORES WITH IMDP RATINGS

| Characteristic | Inclusive sample (n = 390) | | | | IMDP after 2 Yr. (n = 334) | | | |
|--|----------------------------|-------------------|-------------------|-------------------|----------------------------|-------------------|-------------------|-------------------|
| | 1 r | 2 SRC | 3 SRC | 4 SRC | 5 r | 6 SRC | 7 SRC | 8 SRC |
| WCS | | | | | | | | |
| Potential for personal challenge and development | .01 | .09 | .09 | .09 | .02 | .10 | .10 | .11 |
| Responsiveness to new demands | -.03 | .03 | .03 | .02 | -.01 | .10 | .11 | .09 |
| Competitiveness desirability (and reward of success) | -.24 ^a | -.30 ^a | -.31 ^a | -.26 ^a | -.24 ^a | -.29 ^a | -.29 ^a | -.27 ^a |
| Tolerance for work pressure | -.04 | .07 | .05 | .08 | -.08 | .00 | -.02 | .00 |
| Conservative security | .11 | .06 | .04 | .03 | .16 | .12 ^a | .11 | .10 |
| Willingness to seek reward in spite of uncertainty | -.09 | .02 | .04 | .02 | -.09 | .03 | .04 | .02 |
| Surround concern | .11 | .07 | .06 | .05 | .13 | .06 | .03 | .02 |
| SCAT | | | | | | | | |
| Verbal score | -.13 | | -.12 ^a | -.17 ^a | -.16 | | -.15 ^a | -.21 ^a |
| Quantative score | -.09 | | -.09 | -.09 | .03 | | .03 | .02 |
| S-Ident | | | | | | | | |
| Leadership | -.13 | | | .02 | -.09 | | | .02 |
| Impulsivity | .11 | | | .10 | .10 | | | .09 |
| Intellectual orientation | .00 | | | .09 | .02 | | | .13 ^a |
| Aloofness | .10 | | | .01 | .06 | | | -.01 |
| Self-depreciation and low morale | .15 | | | .02 | .08 | | | -.05 |
| Lack of tension | -.09 | | | -.03 | -.08 | | | -.08 |
| BSR | | | | | | | | |
| Assertiveness | -.14 | | | -.11 | -.09 | | | -.04 |
| Likeability | -.13 | | | -.07 | -.12 | | | -.05 |
| Intelligence | -.14 | | | .03 | -.15 | | | .00 |
| Emotionality | .06 | | | -.02 | .04 | | | -.03 |
| Responsibility | -.19 | | | -.10 | -.21 | | | -.16 ^a |
| IO | | | | | | | | |
| Independence-autonomy | -.04 | | | .09 | -.04 | | | .09 |
| Social dependency | .04 | | | .05 | .08 | | | .11 |
| Directiveness | -.12 | | | .01 | -.11 | | | -.02 |
| Sociability | -.03 | | | -.07 | .02 | | | -.07 |
| "I tend to be dependent on others." | .15 | | | .11 | .16 | | | .11 |
| Multiple correlation coefficient with IMDP | | .29 | .32 | .41 | | .31 | .34 | .43 |

* These standardized regression coefficients would be statistically significant in a symmetric hypothesis test with $\alpha = .05$.

Thus, the major variables of first-order correlations with IMDP ratings persist when the standardized regression coefficients are examined. The multiple *R* in this examination which involves 25 variables is .43. Roughly 19% of the variance of the IMDP ratings is accounted for. Considering the restrictions placed on the criteria and on the prediction variables as noted earlier, this is substantial prediction.

The increase in *R* which occurs when the 16 personality variables are added to the *WCS* and *SCAT* scores appears to be real rather

than a function of the number of predictors since an *F* test indicates that the difference between *R* = .43 and *R* = .34 is significant at the .05 level.

SUMMARY AND CONCLUSION

First, data based on new college-level employees are reported which indicate the reliability of the *WCS*. The structure of the *WCS* is explored for the samples and suggests that some scores are slightly more correlated with one another than was reported in a prior study.

Second, examination of the new employees for whom all types of data were available suggests that there are no gross differences on motivational bases as measured by the *WCS* between those who remain with the company and those who leave the company at the initiative of the company, or who are rated as having unsatisfactory progress but leave of their own initiative. A statistically significant difference does occur, however, with the employees still with the company higher on *WCS* Score 3, Competitiveness desirability, than those who left at the company's request. A group of *Ss* who left the company at their own initiative and who were rated as showing satisfactory progress at the time appear to be somewhat higher on *WCS* Score 1, Potential for personal challenge and development, and lower on *WCS* Score 5, Conservative security, but the differences are not statistically significant. If such differences, which are the kinds of self-critical expectations that frequently arise in large organizations, were true, projection of the current data would indicate that the differences would still be extremely small. The *SCAT* scores of the employees with ratings of unsatisfactory progress who leave at the initiative of the company are lower than those still with the company. Those who leave the company at their own initiative and with ratings of satisfactory progress have almost identical *SCAT* scores as those still with the company. Since the *SCAT* scores of these latter two groups are the same, and since it was noted subsequently that there is no difference between these two groups on *WCS* Score 3, Competitiveness desirability, the meaningfulness of the differences between the two groups on the other scores, even if they did exist, is brought into question. While it is true that there is a possible small involvement between a high score on *WCS* Score 5, Conservative security, and a poor rating on the criterion, the association is relatively small and appears to fade in the presence of a larger number of variables. At least in these data, then, there is no indication that persons with extraordinary ability or motivation are leaving.

Third, when Initial Management Development Program ratings are used as a criterion, it is clear that the hires who score highest

on *WCS* Score 3, Competitiveness desirability, are those perceived by the company as likely to reach the third level of management the most quickly. It has a larger regression weight than the *SCAT* abilities score although the variance in the latter score is attenuated because no one with *SCAT* scores less than 91 was hired. Also noted was a significant regression coefficient indicating that individuals who score high on *BSR* responsibility are also perceived as moving up quickly. A smaller negative relation between this perception and *S-Ident* intellectual orientation was also noted. Given the results of this study, it apparently is not the "organization man" who does not "rock the boat" who is perceived as likely to be promoted quickly, but the highly motivated, intelligent, and responsible individual who is.

While the sample sizes and the investment in data collection are substantial in the study reported, any practical implications must be advanced quite tentatively. Obviously, the setting is clear for additional studies to amplify and to explore additional aspects of the problem. For example, hard criteria of success and failure will develop not only for the subsample discussed in this analysis, but for the more inclusive sample for which data were collected through the questionnaire. That is, as a natural consequence of the application of the reward systems of the company, better criterion data will become available for analysis of the *WCS*, *SCAT*, and personality scores, and for the understanding of the progress through which people proceed in their careers.

As noted above, however, it is too early to suggest that implications should be drawn seriously from the current research. What is emphasized is that the stability of the *WCS* scores and their apparent involvement in preliminary study with the *IMDP* criterion as reported here suggest need for more detailed study with hard criteria and more substantial samples. Equally, the study suggests that the multiple-factor base of the *WCS* militates for extreme caution in more simplistic and ad hoc theories of work motivation and orientation. Finally, the persistent relationship of *SCAT* to a criterion like *IMDP* suggests the importance of general abilities in selection. Hopefully, further research on motiva-

tional factors will lead to equally consistent results.

REFERENCES

- BORGATTA, E. F. A very short test of personality: The Behavioral Self-Rating (BSR) form. *Psychological Reports*, 1964, 14, 275-284.
- BORGATTA, E. F. A short test of personality: The S-Ident form. *Journal of Educational Research*, 1965, 63, 309-317.
- BORGATTA, E. F. The Work Components Study: A set of measures for work motivation. *Journal of Psychological Studies*, 1967, 16, 1-11.
- BORGATTA, E. F., & BOHRNSTEDT, G. W. A short-form personality inventory: The Interpersonal Orientations (IO) form. *Journal of Experimental Education*, 1968, 36, 1-6.
- BORGATTA, E. F., FORD, R. N., & BOHRNSTEDT, G. W. The Work Components Study (WCS): A revised set of measures for work motivation. *Multivariate Behavioral Research*, 1968, 3, 403-414.
- CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, 16, 297-334.
- CRONBACH, L. J. *Essentials of Psychological Testing* (2nd ed.). New York: Harper and Row, 1960.
- EDUCATIONAL TESTING SERVICE. *Cooperative School and College Ability Tests*. Princeton: Educational Testing Service, 1964.
- HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. *The motivation to work*. New York: Wiley, 1959.

(Received July 31, 1968)

INFLUENCE OF SEX ROLES ON THE MANIFESTATION OF LEADERSHIP¹

EDWIN I. MEGARGEE²

*Florida State University*³

Situational factors influencing the manifestation of dominance (*Do*) were investigated by pairing Ss high and low on the CPI *Do* scale and having them interact in tasks in which one had to lead and the other follow. In experiments using a masculine industrial task and a sexually neutral clerical task, the following *S* pairs were studied: High and Low *Do* men (Group 1), High *Do* men and Low *Do* women (Group 2), High *Do* women and Low *Do* men (Group 3), High and Low *Do* women (Group 4). Assumption of leadership by the High *Do* women in Group 3 was significantly lower in both studies. This was attributed to sex role conflict inhibiting the manifestation of *Do*. Analyses of the decision-making process supported this interpretation.

In a recent study in this journal, Megargee, Bogart, and Anderson (1966) reported the results of an attempt to predict leadership in a simulated industrial task using the Dominance scale of the CPI. They found that when men who were high in dominance (High *Do*) were paired with men who were low in dominance (Low *Do*) and exposed to a situation in which one had to act as the leader and the other as the follower, the High *Do* individual assumed the leader role 90% of the time when the instructions stressed leadership, but only 56% of the time when the instructions did not stress leadership. They concluded that "the conditions under which leadership is to be exercised are as important as the personality trait of dominance in determining whether or not dominant behavior will be manifested [p. 295]." They suggested that further research should be undertaken in an effort to determine the situational factors which facilitate or inhibit the overt expression of the trait of dominance.

In addition to its intrinsic theoretical interest, this problem is particularly important

for the applied psychologist. Knowing how personality traits interact with social situations will enable him to predict behavior more accurately. Moreover such data would help him design settings which would allow the High *Do* individual to express his leadership ability to its fullest. Research by Smelser (1961) has shown that when two people must cooperate on a task, the productivity or achievement of the group is highest when the High *Do* individual is the leader and the Low *Do* individual the follower. If certain factors inhibiting the High *Do* individual from assuming the leader role can be identified, then steps might be taken which would remove these impediments, thereby increasing the achievement level of the group and the job satisfaction of its members.

The studies to be reported in the present paper investigated how social sex role prescriptions influence the expression of leadership by High *Do* men and women. In our society it is generally considered appropriate for men to dominate women but not vice versa. Most managerial or executive positions are held by men, and while women do not usually feel uncomfortable working for men, men may feel quite discomfited working at the direction of women. It seemed likely that these social role prescriptions would act to inhibit High *Do* women from assuming leadership when paired with Low *Do* men, but might facilitate the assumption of leadership by High *Do* men paired with Low *Do* women.

¹ The writer gratefully acknowledges the assistance of Lynn Godbout, who served as *E* in Study I, and of Patricia Freeman, who served as *E* in both studies. Susan Hamner assisted in this research and Jack Hokanson kindly commented on an earlier draft of this manuscript.

² Requests for reprints should be sent to the author, Department of Psychology, Florida State University, Tallahassee, Florida 32306.

³ This research was performed while the author was at the University of Texas, Austin.

STUDY I

Method

Subjects. A 113-item test labelled the Gough Inventory which consisted of all the items from the *CPI Do*, *Cm*, and *Gi* scales was administered to approximately 600 students in introductory psychology classes at the University of Texas, Austin. From this pool of *Ss*, four groups were formed with 20 pairs of *Ss* in each. Group 1 consisted of High *Do* men paired with Low *Do* men, Group 2 of High *Do* men paired with Low *Do* women, Group 3 of High *Do* women and Low *Do* men, and Group 4 of High *Do* women and Low *Do* women. The *Ss* in each pair were at least 20 *T* score points apart on the *Do* scale.

Apparatus. In their original study, Megargee et al. (1966) used a large box which rested on its side in such a way that the follower had to crawl into it on his hands and knees. While this menial position was well suited to an all-male study such as theirs, it was necessary to modify the apparatus somewhat before it could be used by women wearing skirts and stockings. Therefore, the box was placed in an upright position so that it resembled a large telephone booth without a door. Midway up the side opposite the entrance, 100 $\frac{3}{8}$ -in. holes were drilled 2 in. apart in a 10 \times 10 square pattern. Each hole was filled with a slot-headed bolt 1 in. long and $\frac{1}{4}$ in. in diameter with the slotted head and a washer inside the box, and a square nut tightly screwed onto the bolt outside the box. Because of the narrowness of the $\frac{1}{4}$ -in. bolt relative to the $\frac{3}{8}$ -in. hole, the only way the nut on the outside could be unscrewed efficiently was for one partner to enter the box and hold the bolt in place with a screwdriver while the other partner remained outside and unscrewed the nut with a wrench. The size of the box precluded one person manipulating both the wrench and the screwdriver simultaneously.

Five of the nuts were painted red, 20 were painted yellow, 25 were painted green, and 50 were unpainted. The colors were randomly distributed around the grid. None of the bolts or washers inside the

box were painted, so the person on the inside had no way of determining the location of the different colored nuts.

Procedure. After the pairs of *Ss* had been formed, the individuals were contacted by phone and a time arranged when both could come to the laboratory. At the appointed time each pair was led to the room in which the apparatus was set up. The following instructions were then read with the italicized words emphasized:

This is a study of the relation between the Gough Inventory and *leadership* under stress. This box represents a machine and you are a team of troubleshooters who are to repair it in the fastest possible time. The repair that must be made is to remove all the yellow nuts, leaving the red, green, and unpainted ones in place. One person, who is the *leader*, is to stay outside in front of the machine and the other, who is the *follower*, must go inside. The *leader* must locate the yellow nuts, call out their location to the *follower*, and remove them using this wrench. The *follower* must *obey* the *leader's commands* and, using this screwdriver, hold the bolts in place while the *leader* removes the nuts. It is up to you to decide who will be the *leader* and who will be the *follower*.

Any questions? OK. I shall start timing you now.

Results

The results are presented in the upper portion of Table 1. In Groups 1 and 4, in which both partners were of the same sex, 75% of the High *Do* men and 70% of the High *Do* women took the leader role. This replicated the findings on the predictive validity of the *Do* scale for men obtained by Megargee, Bogart, and Anderson (1966) and extended them to women.

In Group 2, High *Do* men were paired with Low *Do* women. With differences in dominance and social role expectations both operating in the same direction, the manifestation of leadership by the High *Do* men was facilitated and 90% assumed the leader role. In Group 3, in which High *Do* women were paired with Low *Do* men, dominance conflicted with sex role. As expected, this inhibited the assumption of leadership by the High *Do* women. Only 20% assumed the leader role over the Low *Do* men. These differences between the four groups were highly significant ($\chi^2 = 23.96, p < .001$).

Discussion

These results clearly indicated that social role conflict could seriously inhibit the ex-

TABLE 1

NUMBER OF HIGH *Do* *Ss* ASSUMING LEADER AND FOLLOWER ROLES IN THE TWO STUDIES

| Study I ^a | Group | | | |
|----------------------------|-------|----|----|----|
| | 1 | 2 | 3 | 4 |
| No. assuming leader role | 15 | 18 | 4 | 14 |
| No. assuming follower role | 5 | 2 | 16 | 6 |
| <i>N</i> | 20 | 20 | 20 | 20 |
| Study II ^b | | | | |
| No. assuming leader role | 10 | 14 | 4 | 12 |
| No. assuming follower role | 6 | 2 | 12 | 4 |
| <i>N</i> | 16 | 16 | 16 | 16 |

^a $\chi^2 = 23.96, p < .001$.

^b $\chi^2 = 14.94, p < .005$.

pression of leadership by High *Do* women. This raised two questions:

1. How generalizeable were these findings? The simulated industrial task used in this study was an extremely masculine one on which it would be natural for women to defer to men. Could similar results be obtained on a sexually neutral task?

2. Were the differences observed in Group 3 the result of increased assertiveness by the Low *Do* men or of greater submissiveness on the part of the High *Do* women?

To answer these questions a second study was designed and carried out.

STUDY II

Method

Subjects. Students enrolled in introductory psychology classes at the University of Texas, Austin, were pretested with the Gough Inventory and assigned to groups in the same fashion as in Study I except that 16 rather than 20 pairs of Ss were used in each group.

Apparatus. The purpose of this study was to replicate Study I using a sexually neutral two-person task with well-defined leader and follower roles and instructions which emphasized leadership. After some thought it was decided to use a simulated clerical situation with the leader dictating to the follower. The setting was designed to emphasize the difference between the leader and follower roles. It consisted of a table with a screen in the middle and chairs on either side. The leader's chair was an executive-type swivel chair with arms. In front of it rested a leather-covered loose-leaf binder containing the material to be dictated. A sign on the screen above the binder read "LEADER'S SIDE." The follower's chair was a straight-backed wooden chair. He had no binder for his papers and on his side of the screen a sign read "FOLLOWER'S SIDE."

The leader's notebook contained a page from the Stroop Color-Word Test (Stroop, 1935) which had the names of four colors, each printed in ink of a different color, such as the word "red" printed in blue ink. The follower was supplied with a mimeographed form on which the initial letters of the four color names were printed, once for each word on the leaders' sheet.

Procedure. Each pair of Ss was led to the testing room where the following instructions were read with the italicized words emphasized:

This is a study of the relation between the Gough Inventory and *leadership* under stress. One important aspect of *leadership* is the ability to concentrate, remain calm, and accurately give directions to a subordinate. This is what is required by this task. The *leader* will have to absorb information and rapidly pass it on to the *follower*.

The information to be transmitted is the color of the ink on a series of words. This is the *leader's* side of the table. On the table is a sheet of paper on which are printed the names of four colors: blue, green, orange, and red, as you can see in the *leader's* sample. Four different colors of ink, blue, green, orange, and red, have been used in printing the color names. As you can see, the word "orange" is printed in blue ink. The job of the *leader* is to tell the *follower* the color ink used, ignoring the printed word. Since the first word is printed in blue ink, the *leader* should call "Blue"; since the second word is printed in orange ink, he should call out "Orange," and so on.

[*E* moves around table.]

This is the *follower's* side of the table. The *follower* must record the information given him by the *leader*. The *follower's* sheet of paper has the letters "B," "G," "O," and "R" printed, once for each word on the *leader's* sheet. As the *leader* calls out the colors of the ink, the *follower* must record the information by crossing out the letter corresponding to the color, B for blue, G for green, O for orange, and R for red.

This sample has been filled out to correspond to the *leader's* sample. Remember the first word on the *leader's* sample was "Orange" printed in blue ink; therefore the *leader* should have said "Blue" and the *follower* should cross out the letter "b." [If not clear give more examples.]

Thus the *leader* must transmit as much information to his subordinate as possible in the time allowed, as if giving orders in a crisis. The *follower* must record what his *leader* tells him.

It is up to you to decide who will be the *leader* and who will be the *follower*. When you have decided, take your seats at the *leader* or *follower's* sides of the table. When the *leader* is sure the *follower* is ready to record the information, he should call out "Start," turn over the sample page, and start calling out the names of the different-colored inks. When the *leader* calls out "Start," I shall begin timing you; the *leader* will have 90 sec. to transmit as much information as possible.

Any questions?

OK. Decide who will be *leader* and who will be *follower* and take your positions.

After reading the instructions, *E* turned on a tape recorder which recorded the discussions which preceded the choice of leader. Later *E* transcribed these records, noting any additional nonverbal behavior such as an *S* simply sitting down in one position or shrugging when asked his opinion.

Results and Discussion

The data regarding leadership choice are presented in the lower section of Table 1. Once again significant differences were obtained ($\chi^2 = 14.94$, $p < .005$). The proportion of High *Do* Ss in each group assuming

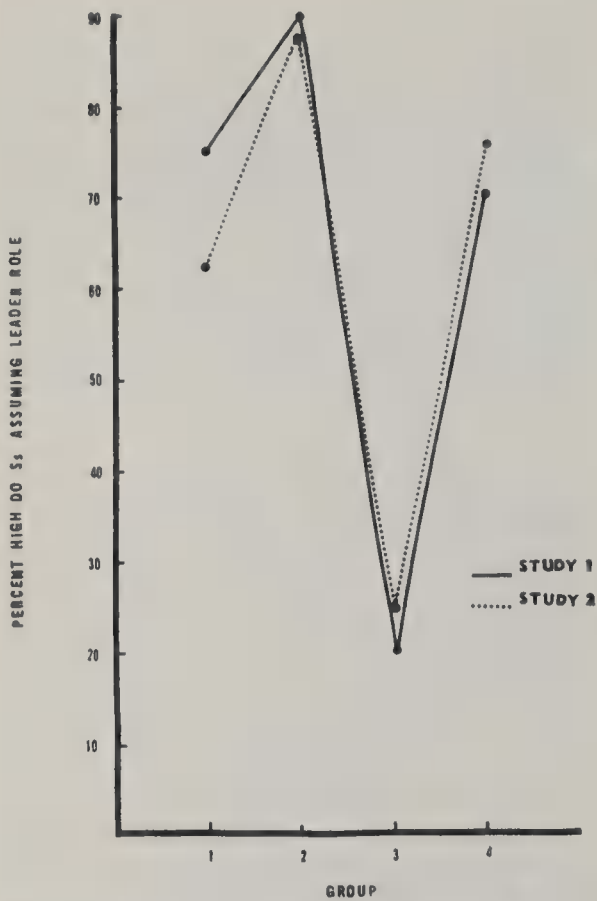


FIG. 1. Proportion of High *Do* Ss assuming leadership in Studies I and II.

the leader role in the two studies are plotted together in Figure 1. It can be seen that the two studies yielded virtually identical results. The decrease in the manifestation of leadership by High *Do* women when paired with Low *Do* men is therefore not limited to highly masculine tasks such as that used in Study I.

Next, the records and notes of the decision-making process were analyzed to determine whether this phenomenon was the result of increased submissiveness on the part of the High *Do* women or of greater assertiveness on the part of the Low *Do* men. It was first noted which *S* made the final decision. Next the behavior of the partner who made the decision was studied to determine whether he appointed himself or his partner leader. Those who appointed themselves leaders were further divided into those who said in effect, "I'll be leader," and those who said, "You be the follower." Similarly those who appointed the other partner leader were subdivided into

those who said, "You be leader," and those who said, "I'll be follower." This third level of analysis did not prove to be useful in differentiating the groups. The behavior of those who allowed their partners to decide was also examined and a record kept of those who actively said, "You decide," and those who simply remained acquiescent or said nothing. This, too, did not prove enlightening.

Interesting group differences were found in the analysis of which partner made the final decision and, more important, whether he decided he or his partner should be leader. Examination of these data in Table 2 shows that Groups 1, 2, and 4, in which dominance did not conflict with social role, followed quite similar patterns. The first row in Table 2 shows the proportion of High *Do* Ss in each group who actually made the final decision. It can be seen that the High *Do* Ss in these three groups often let the Low *Do* partner make the decision. The major difference between the High and Low *Do* Ss was not in who made the decision, but in the nature of the decision which was made. The second row in Table 2 shows that when the High *Do* S in these three groups made the decision, he usually appointed himself leader. However, when the Low *Do* S in these groups made the decision, he generally appointed his partner leader. (In many of these cases the High *Do* S had indicated verbally or nonverbally that while the decision was up to the Low *Do* partner, the High *Do* S would not be averse to assuming leadership.)

In Group 3, in which the pairing of High *Do* women with Low *Do* men brought dominance and sex role into conflict, there was a major difference in the behavior of the High *Do* Ss but not in the behavior of their Low *Do* partners. The High *Do* women made the final decision more often than the High *Do* Ss in any other group, and 91% of the time they appointed their Low *Do* male partners as leader. This was in marked contrast to the behavior of the High *Do* Ss in other groups who, if they made the final decision, never selected their partner as leader more than 33% of the time.

On the other hand, a comparison of the behavior of the Low *Do* men in Group 3 with that of the Low *Do* Ss in the other three

TABLE 2

NUMBER OF HI AND LO *Do* Ss IN EACH GROUP MAKING FINAL DECISION AND
THE NATURE OF THE DECISION MADE

| Ss | | | Group | | | |
|--|--------------|----------|-------|----|----|----|
| | | | 1 | 2 | 3 | 4 |
| No. of Ss making decision | Hi <i>Do</i> | <i>n</i> | 9 | 8 | 11 | 6 |
| | | % | 56 | 50 | 69 | 38 |
| | Lo <i>Do</i> | <i>n</i> | 7 | 8 | 5 | 10 |
| | | % | 44 | 50 | 31 | 62 |
| No. of Ss appointing self leader | Hi <i>Do</i> | <i>n</i> | 6 | 7 | 1 | 4 |
| | | % | 67 | 88 | 09 | 67 |
| | Lo <i>Do</i> | <i>n</i> | 3 | 1 | 2 | 2 |
| | | % | 43 | 12 | 40 | 20 |
| No. of Ss appointing partner leader | Hi <i>Do</i> | <i>n</i> | 3 | 1 | 10 | 2 |
| | | % | 33 | 12 | 91 | 33 |
| | Lo <i>Do</i> | <i>n</i> | 4 | 7 | 3 | 8 |
| | | % | 57 | 88 | 60 | 80 |

groups shows no difference. When the Low *Do* men in Group 3 made the final decision, they appointed themselves as leader 40% of the time and their High *Do* women partners as leader 60% of the time. This was almost identical to Group 1 in which the Low *Do* men appointed themselves as leader 43% of the time and their High *Do* male partners 57% of the time.

This analysis of the decision-making process thus indicated that the low incidence of High *Do* women leaders in Group 3 was not the result of greater assertiveness by the Low *Do* men but instead of reluctance by the High *Do* women to assume overt leadership over a male partner.

The first implication of these two studies is that when predicting leadership, counselors should consider not only dominance and the saliency of leadership in the situation (Megargee et al., 1966), but also the effect of social roles on the overt expression of dominance.

Recent research by Fenelon (1966) has demonstrated that further research is necessary to determine which social situations inhibit or facilitate the expression of dominance. Fenelon studied pairs of High and Low *Do* white and Negro coeds using the clerical task described in Study II. In the biracial situation, Fenelon found that, contrary to expectation, the Negroes took the leader role twice as often as the white girls no matter what the relative *Do* scores. Further research is currently underway by Carl Rubinroit to determine if similar patterns are to be found among Anglo, Negro, and Mexican-American boys in a thoroughly integrated lower class high school.

One might also infer from the present data and from studies such as that of Smelser (1961) that the reluctance of High *Do* women to assert leadership over Low *Do* men would result in less productivity and job satisfaction on the part of both partners. However, re-

search to date has generally overlooked heterosexual work groups, and such studies must be done before this conclusion can be reached validly. Investigations comparing the effectiveness of groups made up of High and Low *Do* men and women in both leader and follower roles should employ both situations in which Ss select the leader and groups in which the leader and follower roles are assigned. It is quite possible that the High *Do* woman might function more effectively when she is appointed leader than when she must assume the leader role on her own initiative. If such research demonstrates that heterosexual groups function better when the High *Do* partner is the leader, then further studies should be undertaken to determine ways in

which the reluctance of the High *Do* woman to assume leadership can be overcome.

REFERENCES

- FENELON, J. R. The influence of race on leadership prediction. Unpublished master's thesis, University of Texas, Austin, 1966.
- MEGARGEE, E. I., BOGART, P., & ANDERSON, B. J. The prediction of leadership in a simulated industrial task. *Journal of Applied Psychology*, 1966, **50**, 292-295.
- SMELSER, W. T. Dominance as a factor in achievement and perception in cooperative problem-solving interactions. *Journal of Abnormal and Social Psychology*, 1961, **62**, 535-542.
- STROOP, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, **18**, 643-662.

(Received August 2, 1968)

THE GRE PSYCHOLOGY TEST AS AN UNOBTRUSIVE MEASURE OF MOTIVATION

ROBERT B. EWEN¹

New York University

Failure to prepare for the Psychology Achievement Test of the Graduate Record Examination (GRE-P) may indicate low motivation. Therefore, GRE-P may serve as an unobtrusive measure of motivation necessary for success in graduate school whether or not the content of the test taps abilities necessary for success. To test this hypothesis, records of 31 males enrolled in various graduate psychology programs at NYU dating from 1960 were obtained. Predictors included GRE-P, GRE Verbal and Quantitative Aptitude Tests, Miller Analogies Test (MAT), undergraduate overall and undergraduate psychology grade-point average (GPA), and number of psychology courses taken prior to the GRE. Criteria included percentage of "A" grades in graduate school and graduation versus termination. Only GRE-P and a difference score consisting of GRE-P minus MAT showed significant validity against the criteria. The results were interpreted as supporting the hypothesis.

The high cost of errors in the selection of students for admission to graduate school in psychology underlines the need for valid selection procedures. As is the case in any selection situation, acceptance of an applicant who fails to succeed involves considerable waste in time and expense on the part of both the institution and the applicant, while rejection of an applicant who would have succeeded deprives the institution of a useful member and the applicant of a position which he merits. The latter type of error is difficult to research because the necessary conditions, wherein all candidates are accepted regardless of test scores, are rarely found in practice; this paper will therefore be devoted to the problem of acceptance of unsuccessful candidates.

A recent monograph by Webb, Campbell, Schwartz, and Sechrest (1966) deals with the use in social science research of unobtrusive measures, where data are not obtained by interview or questionnaire. The authors present examples both from fiction and from real life:

The singular Sherlock Holmes had been reunited with his friend, Dr. Watson, . . . and both walked to Watson's newly acquired office. The practice was located in a duplex of two physician's suites, both of which had been for sale. . . . Holmes summarily told Watson that he had made a wise choice in purchasing the practice that he did, rather than the one on the other side of the duplex. The data? The steps were more worn on Watson's side than on his competitor's [p. 35].

A Chicago automobile dealer, Z. Frank, estimates the popularity of different radio stations by having mechanics record the position of the dial in all cars brought in for service. . . . These data are then used to select radio stations to carry the dealer's advertising [p. 39].

Webb et al. (1966) raise cogent criticisms concerning the frequent use of interviews and questionnaires in social science research:

We lament this overdependence upon a single, fallible method. Interviews and questionnaires intrude as a foreign element into the social setting they would describe, they create as well as measure attitudes, they elicit atypical roles and responses, they are limited to those who are accessible and will cooperate, and the responses obtained are produced in part by dimensions of individual differences irrelevant to the topic at hand. *But the principal objection is that they are used alone* [p. 1].

An objective test would not ordinarily be considered an unobtrusive measure; in fact, it is usually just the opposite. In exceptional instances, however, a written test may act as an unobtrusive measure. The Michigan Vocabulary Profile Test measures vocabulary

¹ The author is indebted to Joseph Weitz for his assistance throughout the course of this study, and to Abraham K. Korman for comments and suggestions. Requests for reprints should be sent to the author, Department of Psychology, New York University, 21 Washington Place, Third Floor, New York, New York 10003.

level in human relations, commerce, government, physical sciences, biological sciences, mathematics, fine arts, and sports. However, scores of the various subparts are not highly intercorrelated, indicating the absence of a general verbal comprehension factor. Therefore, Guion (1965, p. 313) has classified this test as a disguised interest test which reflects the fact that an individual is more likely to learn the jargon of fields of activity in which he is interested, rather than as a test of verbal ability.

A consideration of GRE-P suggests that it may well serve as an unobtrusive measure of the motivation and interest of the applicant. The GRE-P is an achievement test which consists of objective multiple-choice items dealing with various areas of psychology, and is intended for selection at the graduate school level. Students know that this examination is not required by all universities, and hence an applicant to a university which does require this test is likely to think that the university regards the result as important (whether or not this is in fact true).

Let us don the guise of Sherlock Holmes and attempt to deduce the behavior of students with high ability and varying degrees of motivation.² A highly motivated and able student is likely to reason that this test will have at least some effect on his chances of entering the graduate school of his choice, and should therefore take steps to prepare himself for it. Since GRE-P is an achievement test for which studying is likely to have an effect, and since the student is assumed to be an able one, preparation should result in a high score. In graduate school, the combination of high motivation and high ability should produce success. Therefore, high GRE-P scores will be related to success in graduate school through the underlying variable of motivation.

A student with high ability but low motivation, however, is unlikely to engage in behavior involving preparation for GRE-P. The GRE-P is sufficiently comprehensive so that lack of preparation will prohibit a high score for many Ss, despite high ability. In

graduate school, where motivation as well as ability is necessary to produce "A" grades and an acceptable dissertation, the lack of motivation is likely to have a debilitating effect on performance that may well not have been apparent at lower educational levels. Consequently, low GRE-P scores will be related to lack of success in graduate school.

Thus, whether or not GRE-P is a valid predictor in the ordinary sense of tapping abilities necessary for success in graduate school, it may well be useful as an unobtrusive measure of motivation. The present study was designed as a first step towards evaluating this hypothesis.

METHOD

Subjects

Records of 31 males enrolled in the graduate psychology program at New York University no earlier than the fall of 1960 were obtained. Since sex differences might well be operating in this situation, it was decided to reserve an investigation of female Ss for a subsequent study. Including students from far in the past was inadvisable because this would involve dealing with selection standards that were out of date and no longer of interest; the year 1960 was selected as the cutoff point because it was the most recent point in time that would yield a sample of 30 or more Ss. Since the focal point of the study was GRE-P, any student enrolled in 1960 or later but for whom GRE scores (or criterion scores) were missing was excluded from the study.

Sixteen Ss successfully completed the program and received the PhD degree; 15 were dropped by NYU or withdrew without either receiving this degree or transferring to a graduate psychology program at another university with a favorable recommendation from NYU. At the time in question, there were four graduate psychology programs at NYU: social, experimental, industrial, and clinical. Of the graduates, 1 was in the social program, 6 were in experimental, 3 were in industrial, and 6 were in clinical. Of the terminators, 2 were in social, 6 were in experimental, 5 were in industrial, and 2 were in clinical. This sample somewhat relatively underrepresents the social and clinical programs and overrepresents the experimental program (the termination ratios should not be presumed to reflect the norm of any program), but it does at least include some cross section of the four programs at NYU.

Instruments

The predictors used in this study were the Verbal (GRE-V), Quantitative (GRE-Q), and Psychology (GRE-P) parts of the Graduate Record Examination; the Miller Analogies Test (MAT); overall undergraduate GPA; undergraduate psychology GPA; and the number of courses in psychology taken prior

² Possible alternative explanations will be considered in the Discussion section.

to the GRE. The criteria were graduation versus termination, as described above, and the percentage of "A" grades in graduate school. The latter criterion was based on the assumption that while any one course grade may be inaccurate, a student with a large proportion of A grades is likely to be superior academically to a student with a small proportion of A grades. Percentages were used to control for the fact that some students took fewer courses than others; this was especially true insofar as early terminators were concerned. A grades in three-credit courses were counted as one A, and A grades in six-credit courses were counted as two As, noncredit courses were excluded.

Procedure

Since raw scores on the GRE from different years do not always reflect the same percentile rank, and since some Ss took different forms of the MAT, raw scores on these predictors were not used. Instead, the percentile rank corresponding to the raw score for each S was transformed to a standard score by means of the normal curve table, and this standard score was used as the score for S. Three additional predictors were then developed: an average GRE aptitude score, consisting of standard scores on GRE-V plus GRE-Q divided by two; the difference between standard scores on GRE-P and average GRE aptitude; and the difference between standard scores on GRE-P and MAT. The latter two measures were intended to provide information regarding differences between level of ability and level of achievement.

The mean and standard deviation of each variable, and the intercorrelations among all variables, were then computed. A few Ss were missing data on some of the variables, but in no case was the *N* for any statistic below 27. In addition, partial correlation coefficients were computed between GRE-P and each of the criteria, partialling out the number of psychology courses taken prior to the GRE.

RESULTS

The means and standard deviations of the predictors and criteria are shown in Table 1. The high means and small standard deviations for GRE-V and GRE-Q indicate that restriction of range was operating on these variables; the sample is clearly one of high ability insofar as these two tests are concerned. Restriction of range was less apparent on GRE-P, however, and did not appear to be operating on MAT.

The correlations among the variables are shown in Table 2. For GRE-P, a validity of .44 was obtained against the percentage of A's criterion ($N = 31$, $p < .05$) and a validity of .66 was obtained against the graduation criterion ($N = 31$, $p < .01$). The difference

TABLE 1
MEANS AND STANDARD DEVIATIONS OF
PREDICTORS AND CRITERIA

| Variable | <i>N</i> | \bar{X} | <i>S</i> |
|---|----------|-----------|----------|
| 1 GRE-V | 31 | 1.56 | .75 |
| 2 GRE-Q | 31 | 1.07 | .84 |
| 3 GRE-P | 31 | 0.88 | .62 |
| 4 MAT | 29 | 0.39 | .97 |
| 5 Average GRE aptitude | 31 | 1.32 | .60 |
| 6 Overall GPA ^a | 30 | 2.93 | .41 |
| 7 Psychology GPA ^a | 28 | 3.34 | .49 |
| 8 GRE-P minus average GRE aptitude | 31 | -0.44 | .73 |
| 9 GRE-P minus MAT | 29 | 0.52 | 1.11 |
| 10 Number of previous psy- chology courses | 30 | 7.33 | 3.79 |
| 11 Percentage of "A" grades in graduate school | 31 | 38.16 | 33.64 |
| 12 Graduation ^b | 31 | .52 | .51 |

^a Maximum possible was 4.00.

^b 1 = graduated with a PhD degree; 0 = terminated without receiving PhD degree.

score involving GRE-P minus MAT showed a validity of .49 against the percentage of A's criterion ($N = 29$, $p < .01$) and .43 against the graduation criterion ($N = 29$, $p < .05$). No other correlations between predictors and criteria were statistically significant. The correlation between the two criteria was .68 ($N = 31$, $p < .01$) in the expected direction, with graduates receiving more A's than terminators.

The correlations of $-.68$ between MAT and undergraduate psychology graduate point average and $-.43$ between MAT and overall undergraduate grade-point average are difficult to explain. These findings were sufficiently startling to warrant a reexamination of the raw data, but inspection showed that the indicated negative trends did in fact exist. Since the correlations between MAT and GRE-V, and MAT and average GRE aptitude, were positive and significant in spite of the restriction of range of the GRE aptitude measures, it would seem that if this finding is due to experimental error the fault must lie in the grade-point scores. No control for the quality of undergraduate school was available, and it may be that the GPAs which were obtained at different schools were not comparable to one another and that no great importance should be attached to

TABLE 2
INTERCORRELATIONS AND PREDICTIVE VALIDITIES

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|--|---|-----|-----|------|------------------|-------|--------|-------------------|-------------------|------|-------|-------|
| 1 GRE-V | — | .13 | .23 | .42* | .72 ^a | -.18 | -.35 | -.39 ^a | -.21 | -.10 | -.01 | .17 |
| 2 GRE-Q | | — | .18 | .29 | .78 ^a | -.22 | -.11 | -.49 ^a | -.17 | .17 | .17 | .22 |
| 3 GRE-P | | | — | .07 | .28 | -.03 | .01 | .62 ^a | .50 ^a | .30 | .44* | .66** |
| 4 MAT | | | | — | .47** | -.43* | -.68** | -.32 | -.83 ^a | -.23 | -.28 | -.07 |
| 5 Average GRE aptitude | | | | | — | -.27 | -.30 | -.58 ^a | -.25 | .06 | .12 | .26 |
| 6 Overall GPA | | | | | | — | .69** | .19 | .35 | .27 | .11 | .10 |
| 7 Psychology GPA | | | | | | | — | .24 | .58** | .05 | .28 | .04 |
| 8 GRE-P minus average GRE aptitude | | | | | | | | — | .64 ^a | .20 | .28 | .35 |
| 9 GRE-P minus MAT | | | | | | | | | — | .35 | .49** | .43* |
| 10 Number of previous psychology courses | | | | | | | | | | — | -.02 | .32 |
| 11 Percentage of "A" grades in graduate school | | | | | | | | | | | — | .68** |
| 12 Graduation ^b | | | | | | | | | | | | — |

Note. — Decimal points are omitted.
* Significance was not determined for correlations between a composite score and individual variables appearing in that composite score since the variables in these correlations are not independent.
^b 1 = graduated with a PhD degree; 0 = terminated without receiving a PhD degree.
* $p < .05$.
** $p < .01$.

these results. Shortcomings in grade-point measures have been noted by several writers (e.g., Chansky, 1964). In any case, the size of the negative coefficients is such that the lack of an explanation for these findings is disturbing.

Partialling out the number of psychology courses taken prior to the GRE did not greatly affect the validities of GRE-P. The partial correlation with the percentage of A's criterion was .47, and the partial correlation with the graduation criterion was .62.

DISCUSSION

It would be a serious error to interpret the findings concerning GRE-V and GRE-Q as evidence negative to the predictive usefulness of these tests, because restriction of range undoubtedly prevented any significant validities from being obtained. In fact, this study (albeit inadvertently) is best viewed as the second stage of a multiple-cutoff procedure where Ss have passed the first screening based on ability as shown by the GRE aptitude tests. This is in fact ideal for purposes of testing the present hypothesis, since motivated preparation for GRE-P is likely to lead to a good score only if the student is capable. It should be stressed, however, that the findings of this study cannot be generalized to samples with average or low mean GRE aptitude scores. The apparent absence of restriction

of range on the MAT indicates that selection was in general not based on this test, and the failure of the MAT to correlate with the criteria used in this study does therefore represent a substantive finding.

The major hypothesis investigated by this study, that GRE-P is an unobtrusive measure of motivation, is supported by the significant validities obtained for this test against the two criteria. The result concerning the percentage of A's criterion replicates a finding of Stricker and Huber (1967). These investigators obtained a significant validity of .35 for GRE-P against overall GPA in graduate school, using a sample of 37 students which did not include any terminators and may therefore have resulted in underestimation of the size of the validity coefficient. It is also necessary to show, however, that motivation rather than knowledge of psychology is the intervening variable. There are three findings in the present study that support this contention. First, validities of GRE-P were not appreciably affected when the number of psychology courses taken prior to the GRE was partialled out. Second, correlations between GRE-P and both undergraduate grade-point measures were not significant. As discussed previously, however, it is possible that the grade-point measures were deficient, so this finding should be interpreted with caution. Third, if GRE-P is primarily a measure of knowledge of psychology, it should

be correlated as or more highly with the percentage of A's in graduate school criterion than with the graduation criterion. Instead, the correlations with graduation was numerically higher, and the difference between the two validities approached statistical significance (p between .10 and .05, using the formula for nonindependent r 's [Edwards, 1960, p. 85]). In addition, the significant validities for the difference score involving GRE-P and MAT suggest that factors other than ability are operative. In view of the fact that the sample was a preselected one and that other predictors were not correlated with the criteria used in this study, the validities involving GRE-P are particularly striking.

The results of this study should be interpreted with caution, since the sample was small and since unequivocal identification of motivation as the true intervening variable cannot be claimed. Alternative explanations of the results, perhaps in terms of a social science aptitude tapped by GRE-P, are possible. Also, the present study does not attempt to identify the underlying nature of the motivational variable. The work of Spence, Taylor, and others (e.g., Spence & Farber, 1953; Spence & Taylor, 1951; Taylor, 1951) concerning manifest anxiety and generalized drive level suggests the hypothesis that variations in manifest anxiety better explain variations in the level of preparation for GRE-P.

While further research is necessary to evaluate these possibilities, the present results suggest that for a sample high in verbal and quantitative ability, GRE-P may well serve as an unobtrusive measure of motivation which will significantly improve the prediction of success in graduate school in psychology.

REFERENCES

- CHANSKY, N. M. A note on the grade-point average in research. *Educational and Psychological Measurement*, 1964, **24**, 95-99.
- EDWARDS, A. L. *Experimental design in psychological research*. (rev. ed.) New York: Holt, Rinehart and Winston, 1960.
- GUION, R. M. *Personnel testing*. New York: McGraw-Hill, 1965.
- SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *Journal of Experimental Psychology*, 1953, **45**, 116-119.
- SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *Journal of Experimental Psychology*, 1951, **42**, 183-188.
- STRICKER, G., & HUBER, J. T. The Graduate Record Examination and undergraduate grades as predictors of success in graduate school. *Journal of Educational Research*, 1967, **60**, 466-468.
- TAYLOR, J. A. The relationship of anxiety to the conditioned eyelid response. *Journal of Experimental Psychology*, 1951, **41**, 81-92.
- WEBB, E. J., CAMPBELL, D. T., SCHWARTZ, R. D., & SECHREST, L. *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand-McNally, 1966.

(Received August 14, 1968)

CORRELATES OF JOB SATISFACTION AND JOB DISSATISFACTION AMONG FEMALE CLERICAL WORKERS

L. K. WATERS¹ AND CARRIE WHERRY WATERS

Ohio University

Degree of overall satisfaction, overall dissatisfaction, and overall satisfaction/dissatisfaction were correlated with measures of satisfaction/dissatisfaction with several aspects of the work situation for 160 female clerical workers. Also, tabulations were made of responses to open-ended questions concerning reasons for positive and negative feelings about the company. The results of these analyses offered no support for the two-factor theory of job satisfaction, but were consistent with the traditional framework in which any variable can be both a "satisfier" and a "dissatisfier."

The two-factor theory of job satisfaction (Herzberg, 1966; Herzberg, Mausner, & Snyderman, 1959) proposes that one set of variables in the work situation leads to satisfaction but not dissatisfaction (motivator, satisfier, or intrinsic variables), while another set of variables in the work situation leads to dissatisfaction but not satisfaction (hygienic, dissatisfier, or extrinsic variables). This proposal is in contrast to the more traditional model of job satisfaction in which any work-related variable may contribute to both satisfaction and dissatisfaction.

Several recent studies (Ewen, Smith, Hulin, & Locke, 1966; Graen, 1966; Halpern, 1966; Henrichs & Mischkind, 1967; Wernimont, 1966) have reported data contradictory to the two-factor theory. However, in each of these studies satisfaction/dissatisfaction was measured on a single continuum. If, as the two-factor theory suggests, satisfaction and dissatisfaction are qualitatively different, they should be assessed separately (see Whitsett & Winslow, 1967, for a discussion of this point).

In a more recent article (Hulin & Smith, 1967), measures of satisfaction, dissatisfaction, and satisfaction/dissatisfaction were obtained from different randomly selected groups within the same company. These data are also contradictory to the two-factor theory for both male and female employees.

The present paper reports the relationships of selected job-related variables to separate measures of overall satisfaction, dissatisfaction, and satisfaction/dissatisfaction obtained from the same group of female clerical employees.

METHOD

The respondents in this study were 160 non-supervisory female employees in one regional office of a national insurance company. The employees ranged in age from the late teens to early sixties and all were at least high school graduates. Immediate supervisors of all employees utilized in the study were also females.

A job attitude questionnaire was administered to small groups of employees by the author during a single working day. Respondents were assured that their individual responses would not be made known to the company. Names were requested but employees were given the option of not responding to that item if they "felt uncomfortable" doing so. Approximately 17% did not give their names. Other information (job title, department, etc.) obtained on the questionnaire probably would have been sufficient to identify the respondent.

The job attitude scales were presented in booklet form and consisted of separate overall satisfaction and dissatisfaction scales (always the first two scales, order randomized), an overall satisfaction/dissatisfaction scale, the five scales of the Job Description Index (JDI), and a list of 11 job factors (arranged in alphabetical order) to be rated on a satisfaction/dissatisfaction scale. Ratings of satisfaction/dissatisfaction (both overall and for specific job factors) were made on a 12-point anchored scale and the separate satisfaction and dissatisfaction ratings on 7-point scales which consisted of the appropriate 6 points of the 12-point satisfaction/dissatisfaction scale plus a seventh alternative (not satisfied or not dissatisfied). Immediately after each of the separate satisfaction and dissatisfaction scales,

¹ Requests for reprints should be sent to Lawrence K. Waters, 45 Eden Place, Athens, Ohio.

respondents were asked to indicate "one or two things that most influence your feelings in a positive (or negative) way about your employment at the _____ company." Responses were sorted independently by two judges into motivator, hygienic, wage, or unclassifiable categories. The judges agreed on 89.4% of the responses. Items on which the judges did not agree were labeled unclassifiable. If an employee gave only one job factor as a positive or negative influence, it was given a weight of two; if two job factors were mentioned, they each received a weight of one. If more than two factors were mentioned, only the first two were coded. About 84% of the respondents listed at least one factor as a positive influence, 76% listed at least one job factor as a negative influence, and 8% did not respond to either question.

RESULTS

The mean scores for the overall satisfaction (S), dissatisfaction (D), and satisfaction/dissatisfaction (S/D) scales were 3.58 ($SD = 1.30$), 1.65 ($SD = 1.34$), and 9.02 ($SD = 2.10$), respectively. The correlation between S and D was $-.61$, and the two scales correlated $.78$ and $-.64$ with S/D. To determine if the order in which the S and D scales were presented affected responses to these scales, S, D, and S/D mean ratings were computed separately for employees responding under each of the two presentation orders. The comparisons of mean ratings between the two groups on corresponding scales yielded t 's of less than 1.00 in all three comparisons.

Correlations between the JDI scales and the satisfaction/dissatisfaction rating for the corresponding job factor from the list of job factors were computed to obtain estimates of the consistency of employee responses. These correlations were JDI Work—Work = $.70$, JDI Pay—Salary = $.73$, JDI Promotion—Opportunity for Growth and Advancement = $.63$, JDI Supervision—Competent Supervision = $.80$, JDI Supervision—Considerate Supervision = $.79$, and JDI Co-workers—Co-workers = $.61$. In all cases, the job factor correlated higher with the appropriate JDI scale than with any other job factor from the list, and the correlation between the JDI scale and the corresponding job factor was the largest correlation that the JDI scale had with any job factor.

According to the two-factor model, motivator variables should be related to degree of satisfaction but not degree of dissatisfaction,

TABLE 1
CORRELATIONS BETWEEN JDI SCALES AND
OVERALL JOB SATISFACTION

| JDI area | Overall satisfaction-dissatisfaction | Overall satisfaction | Overall dissatisfaction |
|-------------|--------------------------------------|----------------------|-------------------------|
| Work | .53** | .62** | -.45** |
| Pay | .29** | .33** | -.25** |
| Promotion | .25** | .20* | -.19* |
| Supervision | .41** | .35** | -.32** |
| Co-workers | .26** | .28** | -.13 |

* $p < .05$.

** $p < .01$.

tion, and hygienic variables should be related to dissatisfaction but not satisfaction. No predictions concerning the relationship of motivator and hygienic factor with overall satisfaction/dissatisfaction can be made since the degree of overall satisfaction/dissatisfaction is presumed to be some unspecified composite of positive and negative influences. The traditional model, on the other hand, would predict that any job-related variable may correlate with both satisfaction and dissatisfaction.

The correlations of the five JDI scales and overall S, D, and S/D are shown in Table 1. In general, the relationships of the JDI scales tended to be somewhat larger with S than D, but, except for the JDI Co-worker-D-scale coefficient, all of the correlations were significant and the order of the three more highly related JDI scales was the same for both S and D. One of the problems in interpreting these data in terms of the two-factor framework is the classification of job-related factors as either motivator or hygienic. Ewen et al. (1966) and Hulin and Smith (1967) classified JDI Work and Promotion as motivators and Pay as a hygienic. Supervision and Co-workers were considered ambiguous in terms of classification. Whitsett and Winslow (1967) objected to the use of Promotion as a motivator and Pay as a hygienic (p. 399). Henrichs and Mischkind (1967) deleted pay as a hygienic variable. The only point of agreement seems to be that JDI Work is a motivator. As a motivator, it should be correlated with S but not D according to the

two-factor model. For the female sample of the present study, JDI Work had the highest correlation of the scales with both S, D, and S/D, which is directly opposed to the motivator-hygienic framework. With the exception of Co-workers, the other JDI scales (however classified for the conditions of this study) were significantly related to S, D, and S/D.

The relationships of ratings on different aspects of the job as obtained from the list of job factors to S, D, and S/D are given in Table 2. As was the case for the JDI scales, correlations with S tended to be somewhat larger than with D. With the exception of Co-workers, all of the correlations (for both motivator and hygienic factors) were significant and the pattern of relationships with S and D were very similar ($\rho = .82$). These data support the traditional theory and conflict with the two-factor model.

Although neither the traditional nor the two-factor model make predictions concerning the relative potency of motivator and hygienic factors, several studies have reported generally greater saliency for motivator than hygienic variables (Ewen et al., 1966; Halpern, 1966; Hulin & Smith, 1967; Wernimont, 1966). The mean r 's for both motivator and hygienic clusters were computed (using Fisher's z' transformation) to estimate the relative potency of the two classes of variables. For overall S, D, and S/D scales, motivators were generally more highly

TABLE 2

CORRELATION BETWEEN SATISFACTION/DISSATISFACTION (S/D) WITH DIFFERENT ASPECTS OF THE JOB AND OVERALL SATISFACTION

| Job factors | Overall S/D | Overall S | Overall D |
|--|-------------|-----------|-----------|
| Motivator | | | |
| Opportunity for Growth and Advancement | .50** | .37** | -.36** |
| Responsibility on the Job | .43** | .41** | -.37** |
| Recognition for Work Done | .37** | .40** | -.28** |
| Sense of Achievement | .58** | .51** | -.36** |
| Work | .50** | .51** | -.45** |
| Mean r | .48 | .44 | -.36 |
| Hygienic | | | |
| Competent Supervision | .44** | .44** | -.40** |
| Considerate Supervision | .30** | .33** | -.31** |
| Company Policies & Practices | .29** | .29** | -.16* |
| Co-workers | .13 | .14 | -.06 |
| Physical Working Conditions | .24** | .23** | -.24** |
| Mean r | .28 | .29 | -.24 |
| Salary | .47** | .43** | -.28** |

* $p < .05$.
** $p < .01$.

TABLE 3
CLASSIFICATION OF REASONS FOR POSITIVE AND NEGATIVE FEELINGS

| Feelings | Motivator | Hygienic | Wages | Unclassified |
|--------------------------------|-----------|----------|-------|--------------|
| Positive ($\Sigma = 268$) | 16.8% | 70.5% | 5.6% | 7.1% |
| Negative ($\Sigma = 242$) | 15.7% | 49.6% | 20.2% | 14.5% |

related than the hygienic cluster. The data support previous results.

The results from the open-ended questions concerning factors influencing feelings about the company in a positive or negative way are presented in Table 3. The percentage of weighted responses classified as motivators was almost the same for positive and negative influences. However, if wages are excluded from the hygienic category, more hygienic factors were mentioned as positive influences than negative influences. Considering wages as part of the hygienic classification yielded similar percentages for positive and negative influences (76.5 and 70.0, respectively). However, for either definition of the hygienic set, these data are contradictory to the two-factor theory.

DISCUSSION

The results of the correlational analysis and tabulation of responses to the open-ended questions were contradictory to the two-factor theory. In the correlational analysis the patterns of correlations for the various aspects of the job were very similar whether S, D, or S/D was used as the measure of overall attitude about the job. Motivators performed as both satisfiers and dissatisfiers, and hygienic factors were related to both satisfaction and dissatisfaction. In addition, the substantial correlation between overall satisfaction and overall dissatisfaction seems to offer little in support of the contention of the two-factor model that satisfaction and dissatisfaction are qualitatively different. Responses to the open-ended questions indicated that hygienic aspects were mentioned more frequently as reasons for positive feelings than negative feelings.

These results, especially in conjunction with the Hulin and Smith (1967) study, in-

dicates that even when satisfaction and dissatisfaction are assessed separately as suggested by the proponents of the two-factor model, no support is found for the theory for female employees. Both of these studies agree in conclusion with previous research (Ewen et al., 1966; Graen, 1966; Halpern, 1966; Henrichs & Mischkind, 1967; Wernimont, 1966) using a single overall satisfaction/dissatisfaction scale. The weight of these studies suggests that the results supporting the two-factor model are method-bound and that the model offers little to the understanding of worker attitudes.

REFERENCES

- EWEN, R. B., SMITH, P. C., HULIN, C. L., & LOCKE, E. A. An empirical test of the Herzberg two-factor theory. *Journal of Applied Psychology*, 1966, **50**, 544-550.
- GRAEN, G. B. Addendum to "An empirical test of the Herzberg two-factor theory." *Journal of Applied Psychology*, 1966, **50**, 551-555.
- HALPERN, G. Relative contributions of motivator and hygiene factors to overall job satisfaction. *Journal of Applied Psychology*, 1966, **50**, 198-200.
- HERZBERG, F. *Work and the nature of man*. New York: World, 1966.
- HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. *The motivation to work*. New York: Wiley, 1959.
- HENRICH, J. R., & MISCHKIND, L. A. Empirical and theoretical limitations of the two-factor hypothesis of job satisfaction. *Journal of Applied Psychology*, 1967, **51**, 191-200.
- HULIN, C. L., & SMITH, P. A. An empirical investigation of two implications of the two-factor theory of job satisfaction. *Journal of Applied Psychology*, 1967, **51**, 396-402.
- WERNIMONT, P. F. Intrinsic and extrinsic factors in job satisfaction. *Journal of Applied Psychology*, 1966, **50**, 41-50.
- WHITSETT, D. A., & WINSLOW, E. K. An analysis of studies critical of the motivator-hygiene theory. *Personnel Psychology*, 1967, **20**, 391-415.

(Received August 19, 1968)

CREATIVITY AND ACADEMIC MAJOR: BUSINESS VERSUS ENGLISH MAJORS¹

RUSSELL EISENMAN²

Temple University

Business and English majors were compared on two measures of creativity, unusual uses for common objects and the Personal Opinion Survey, in order to explore further the interpretation of Maier and Hoffman (1961) regarding the inhibiting organizational effects on creativity. Chi-square tests revealed that English majors were superior to business majors on both creativity tests (p 's < .01) suggesting that a selective factor is operative: English attracts students high in creativity while business attracts students low in creativity. These results make the Maier and Hoffman emphasis on organizational effects somewhat dubious, since their results can be explained on the basis of this selective factor. Further comparison of the 48 Ss with 229 Ss previously tested indicated that business majors were significantly low on creativity while English majors were significantly high.

Maier and Hoffman (1961) reported an investigation of creativity in groups varying in the amount of experience and identification with existing organizations. Their expressed purpose was to consider the effect of experience and identification with business and industrial organizations on creativity. Employing a role-playing case, the Change of Work Procedure problem (Maier, 1955), the authors studied creative problem solving in four groups: (a) an employed group, consisting of industrial foremen, airline managers, training directors, hospital managers, and nursing supervisors; (b) business administration students; (c) psychology of human relations students; and (d) introductory psychology students. The results clearly indicated that the greater the experience and identification with business or industry the lower the creativity. Maier and Hoffman (1961) concluded that "The results of this study provide suggestive empirical support for the proposition that the usual formal authority structure found in present-day organizations tends to inhibit the expression of the creative potential of their members [p. 279]."

Since the production of integrative—creative—solutions by groups with little or no identification and experience in business was

three times as large as the proportion of integrative solutions by Ss more identified with business or industrial organizations, there seems little room for doubt that Ss were different in their performance on the assigned task. To the extent that the role-playing task effectively measures creativity, the business-industrial Ss performed markedly lower in creativity. However, in spite of the great difference between business versus nonbusiness Ss there is one major conceptual problem, which leaves the basis of the results in question. Maier and Hoffman (1961) favor the interpretation that the formal authority system inhibits creativity by promoting an atmosphere in which employees simply wish to do what the boss considers the right thing to do. While their results are consistent with such an explanation, there is yet another possible interpretation which needs to be considered. Could it not be that there is a selective factor operating, such that business and industry tend to attract Ss who are relatively lower in creativity than persons in other fields? If this selectivity were accepted as the explanation of the Maier and Hoffman results, it would be unnecessary to explain the findings in terms of length of association and identification with business or industry. Instead, the major explanatory power would be carried by the concept of initial differences among the groups tested, such that business majors in college are relatively low in crea-

¹ This research was aided by a grant from the Trustees of Temple University.

² Requests for reprints should be sent to the author, Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.

tivity even before their employment in business or industrial organizations.

The major difference between the possible explanation advanced here and the one favored by Maier and Hoffman (1961) is that their explanation emphasizes the inhibiting effect of organizational authority, while the current hypothesis points out that organizational demands need not be called upon to explain their results. Instead, individual differences may be present in Ss in various fields. The specific hypothesis of the present study was that business students would score lower on two creativity tests relative to English majors. The two creativity tests employed were an unusual uses test in which S is required to give uses for common objects, with the higher scores going to Ss who produce original (statistically infrequent) responses (Eisenman & Robinson, 1968); and the Personal Opinion Survey (Eisenman, 1968), a 30-item paper-and-pencil personality measure of creativity. English majors were chosen because it is of interest to see if their alleged interest in literary originality would be associated with a tendency to accept the unreal and the unusual. If such a tendency exists among English majors it should increase their creativity, since creativity is frequently conceptualized as an ability to ignore the obvious and consider unusual ways of doing things (Barron, 1963; Taylor, 1964).

METHOD

Subjects

The Ss were 48 students at Temple University, including 20 English majors and 28 business majors. These Ss constituted all the business and English majors from two classes with a total of 65 students, one a business course and one a course in English. The Ss were tested by an assistant during their class, with the permission of their teachers.

Measures

Unusual uses is a common creativity measure with different investigators using a similar format, namely, Ss are presented with the names of common objects and asked to list all the uses they can think of for such objects. Originality is defined as statistical infrequency. For example, in the present study any use which appeared less than 5% of the time in the present samples was deemed original. It is also possible to obtain a fluency score simply by adding the number of valid responses, considering as invalid the basic repetition of a theme. "Build a house" and

TABLE 1

NUMBER OF Ss ABOVE AND BELOW THE MEDIAN (*Mdn* = 23.50) ON TOTAL NUMBER OF ORIGINAL USES FOR OBJECTS ON AN UNUSUAL USES TEST

| Classification | Above <i>Mdn</i> | Below <i>Mdn</i> |
|----------------|------------------|------------------|
| Business major | 9 | 19 |
| English major | 15 | 5 |

Note.— $\chi^2 = 6.94$, $df = 1$, $p < .01$.

"build a dog house" would be considered only one response since they both employ the concept of building. However, correlations of fluency and originality are often in the .80's-.90's (Eisenman, 1969; Madaus, 1967) so fluency was not evaluated in the present study.

The Personal Opinion Survey (Eisenman, 1968; Eisenman & Robinson, 1967) is a 30-item, true-false, paper-and-pencil personality measure of creativity. The test is composed of five short-form tests of six items each taken from Child (1965). The subtests are tolerance for complexity, tolerance for ambiguity, scanning, independence of judgment, and regression in the service of the ego. The overall score is used, with a maximum possible score of 30. Higher scores indicate greater creativity. Odd-even reliability has been found to be .86 with the Spearman-Brown prophecy formula.

Since the Personal Opinion Survey was machine scored, interscorer reliability is not an issue. With the unusual uses test, Ss had to list uses for five objects: bricks, spoons, paper clips, paper, and toothpicks. Interscorer agreement ranged from 70-89% for the various objects. Reasonable validity can be claimed for both creativity measures; the reader is referred to Eisenman (1968, 1969).

Although no specific time limit was mentioned, the implied time limit was the 50 min. of the class period.

RESULTS

The performance for business versus English majors is shown in Table 1 for the unusual uses test. It is apparent that on this creativity measure the English majors were significantly superior to the business students.

Table 2 shows the relative performance of

TABLE 2

NUMBER OF Ss ABOVE AND BELOW MEDIAN (*Mdn* = 17.58) ON PERSONAL OPINION SURVEY

| Classification | Above <i>Mdn</i> | Below <i>Mdn</i> |
|----------------|------------------|------------------|
| Business major | 8 | 20 |
| English major | 16 | 4 |

Note.— $\chi^2 = 10.37$, $df = 1$, $p < .01$.

business and English majors on the second creativity measure, the Personal Opinion Survey. Again, the English majors were significantly more creative.

The possibility remained that the results were due to either (a) the normal achievement of business majors on these measures but the markedly high performance of English majors, relative to other samples of Ss; or (b) the normal achievement of English majors but the markedly inferior scores of business majors. As a partial attempt to answer this, the scores of business and English majors were compared with a normative sample of 229 Ss reported in Eisenman (1968). These Ss, composed mainly of college undergraduates, but including a smattering of professional men and laborers, obtained a mean score of 18.87. The business majors' mean of 16.80 is significantly lower via a *t* test ($p < .05$) while the English majors' mean of 20.05 is significantly higher, again by a *t* test with significance beyond the .05 level. Therefore, both (a) and (b) above can be ruled out and the conclusion to be drawn is that, relative to the 229 Ss, the business majors are low in creativity and the English majors are high, as measured by the Personal Opinion Survey.

DISCUSSION

The finding of low creativity in business majors even before they begin employment in business or industrial organizations suggests that for some reason business does not attract very creative students. The implications of this selective factor for the interpretation of Maier and Hoffman's (1961) study would be to suggest that while it is possible that organizations inhibit creativity, with greater length of time in organizations associated with lessened creative performance, the persons who go into these organizations tend, by and large, to be relatively low in creativity. Perhaps the interpretation emphasized here complements the Maier and Hoffman view. If relatively noncreative people are attracted to business as an academic major then they are not likely as leaders to foster creativity on the part of their subordinates.

The higher creativity among English majors is also meaningful, and may serve as a basis

to infer the differences between business and English majors. Barron (1963) has emphasized how rigid control, interest in the concrete and safe, and a generally simple life style are inconsistent with creativity. His creative Ss prefer complexity both in aesthetic preference and in their everyday life, in contrast to the noncreative Ss who achieve whatever prominence they attain by a more conventional adhering to the rules of society. Business tends to place a strong emphasis on the practical, everyday issue of economic competition, while English is a field with more emphasis on fantasy, the "inner life" of man, and less emphasis than business on concrete matters. Conceived in this way, the field of business adopts a stance that is more like that of Barron's low creative Ss, while the field of English shows an interest in areas Barron finds correlated with high creativity.

The results have further implications for student reaction to college climates. It may be predicted that business majors would more readily support the conservative organizations, formal and informal, which oppose student attempts to engender change on the college campus. It might also be supposed that English majors would not be so likely to support campus authority. Whether or not English majors might be overrepresented among students pressing for change would depend additionally on their propensity for action versus their tendency to think rather than to do. An unpublished study of a campus boycott (Eisenman, Aserinsky, & Robinson, 1969) provides modest support for one of the above-mentioned predictions. A large sample of students yielded only five students who actively violated a student boycott of the school cafeteria, thereby supporting the existing authority and opposing the peer group. All five were business majors. While such evidence is admittedly slim, it is entirely consistent with the results obtained in the present study. The present study, like that of Maier and Hoffman (1961), suggests that business does not attract highly creative persons to any great extent.

REFERENCES

- BARRON, F. *Creativity and psychological health*. Princeton, N. J.: Van Nostrand, 1963.

- CHILD, I. L. Personality correlates of esthetic judgment in college students. *Journal of Personality*, 1965, 33, 476-511.
- EISENMAN, R. Values and attitudes in adolescence. In J. F. Adams (Ed.), *Understanding adolescence*. Boston: Allyn & Bacon, 1968.
- EISENMAN, R. Creativity, awareness, and liking. *Journal of Consulting and Clinical Psychology*, 1969, 33, 157-160.
- EISENMAN, R., ASERINSKY, A., & ROBINSON, N. H. Student reaction to social protest: The brown bag boycott. Paper presented at annual meeting of the Southeastern Psychological Association, New Orleans, Feb. 1969.
- EISENMAN, R., & ROBINSON, N. Complexity-simplicity, creativity, intelligence, and other correlates. *Journal of Psychology*, 1967, 67, 331-334.
- EISENMAN, R., & ROBINSON, N. Generality of some complexity-simplicity measures related to creativity. Proceedings of the 76th annual convention of the American Psychological Association, 1968, 3, 141-142.
- MADAUS, G. F. A cross-cultural comparison of the factor structure of selected tests of divergent thinking. *Journal of Social Psychology*, 1967, 73, 13-21.
- MAIER, N. R. F. *Psychology in industry*. (2nd ed.) Boston: Houghton Mifflin, 1955.
- MAIER, N. R. F., & HOFFMAN, L. R. Organization and creative problem solving. *Journal of Applied Psychology*, 1961, 45, 277-280.
- TAYLOR, C. W. (Ed.) *Widening horizons in creativity*. New York: Wiley, 1964.

(Received August 22, 1968)

FORCED-CHOICE PREFERENCES AND RESPONSE LATENCY

BRUCE O. BERGUM¹ AND ROGER P. DOOLEY

Research Laboratories

Xerox Corporation, Rochester, New York

An experiment was conducted to examine the relationship between overt forced-choice preferences and response latencies. A group of 24 males made 66 paired comparisons of 12 stimuli, indicating their preferences on a keyboard and basing their judgments on aesthetic considerations. The results indicated a systematic relationship between overt expressions of preference and the time required to indicate this choice on the keyboard. The most preferred stimuli yielded the shortest latencies, while the least preferred stimuli yielded the longest latencies. These results are in line with those demonstrated in previous studies employing considerably different procedures and conditions. However, the present procedure tended to yield a somewhat more linear relationship than did the previous studies.

A series of recent studies has demonstrated a significant relationship between the judged affective value of a wide range of visual and auditory stimuli and the latency of the judgmental response. In general, this relationship tends to take the form either of an essentially linear function (Bergum & Lehr, 1966; Bergum, Lehr, & Dooley, 1967) or of an inverted J (Bergum & Lehr, 1967; Lehr, Bergum, & Standing, 1966), with positively affective stimuli yielding short latencies in all cases and negatively affective stimuli yielding long latencies. In these earlier studies, however, the failure of negatively affective stimuli to yield consistently longer latencies than neutral stimuli could not be explained readily on the basis of the available data, although a number of hypotheses seemed possible. These included greater affective ambivalence, or uncertainty, toward the neutral stimuli, changes in the affective base line as new stimuli were experienced in the test situation, and the inherent unreliability of single response evaluations. Unfortunately, none of these possibilities could be tested adequately with the procedures employed in the earlier studies.

The purpose of the present experiment was thus to remedy this situation in part by employing a procedure where the basic reliability of the evaluation responses would be more assured, and where shifts in the affective

base line could be controlled. This procedure, paired-comparisons, had the added virtue of providing more precise and reliable estimates of the affective distance between the stimuli in the sample.

METHOD

Subjects

Twenty-four adult, college-educated males served as Ss in this experiment. All were members of the Xerox Research Laboratories staff.

Apparatus

Apparatus for the experiment included a Sawyers 35mm. slide projector, a rear-projection screen, a keyboard and associated event recorder, and an audio tape recorder for presenting instructions.

Materials and Conditions

Six different types of stimuli were employed. These included a typewritten discussion of calendars, a table of salaries for scientists, a line graph, a bar chart, a map showing the distribution of scientists in the United States, and an exploded engineering drawing of an idler assembly. A colored and a black and white version were prepared of each of the six stimulus types and all possible pairings of the 12 stimuli were photographed on 35mm. slides. The position of each stimulus was counterbalanced so that half of the time it appeared in the right half or lower position on the slide. The members of the pairs were labeled "1" or "2" on each slide to indicate their correspondence to the similarly labeled response keys. In all, a total of 66 such slides were prepared.

Each S viewed the 66 stimulus pairings in a different random order to control for possible practice effects and ordering effects among the stimuli.

¹ Requests for reprints should be sent to Bruce O. Bergum, Xerox Corporation, Research Laboratories, 800 Phillips Road, Webster, New York 14580.

The Ss were seated in front of the rear-projection screen and instructed first to view the stimulus pair and then to indicate which member they most preferred by depressing the appropriate response key. They were specifically instructed to employ aesthetic qualities as the basis for judgment to the extent that this was possible, and each S was allowed to proceed at his own pace until all 66 judgments had been made. Response measures included both overt preferences and response latencies.

RESULTS

The data were treated in three different ways. First, the total number of times each stimulus was chosen over all other stimuli was determined across all Ss, and these data converted to percent preferences. Second, the mean preference response latency was determined across all Ss for all comparisons involving each given stimulus. Finally, the 12 stimuli were rank-ordered both in terms of the percent preference results and the latency results and a rank-order correlation performed between these two sets of ranks.

Table 1 lists the 12 stimuli with their associated percent preferences, response latencies, and their associated rank orders. Casual inspection of the two sets of ranks suggests a high degree of relationship between the relative strength of overtly expressed preferences and the response latencies associated with these judgments. The rank-order correlation between these two sets of data is .93, $p < .01$, indicating that both methods are, in fact, measuring very nearly the same thing.

DISCUSSION

The demonstration of an essentially linear relationship between overt preferences and choice-response latencies is in direct support of the results reported earlier in that the more preferred stimuli consistently yielded shorter latencies than did the less preferred stimuli. In the present case, however, the tendency for the neutral, or middle-ground, stimuli to yield the longest latencies was not observed, suggesting that the earlier results may well have been an artifact relating to the testing procedures employed. It seems likely that the longer latencies demonstrated with the earlier procedures resulted from the un-

TABLE 1
PREFERENCES AND LATENCIES

| Stimulus | % preferred | Rank | Latency | Rank |
|---------------|-------------|------|---------|------|
| Map-Color | 89.8 | 1 | 1.03 | 1 |
| Drawing-Color | 80.7 | 2 | 1.13 | 3 |
| Chart-Color | 73.5 | 3 | 1.05 | 2 |
| Graph-Color | 59.8 | 4 | 1.26 | 4 |
| Map-Black | 52.7 | 5 | 1.31 | 5 |
| Chart-Black | 52.3 | 6 | 1.41 | 8 |
| Table-Color | 50.4 | 7 | 1.32 | 6 |
| Text-Color | 38.3 | 8 | 1.42 | 9 |
| Graph-Black | 35.2 | 9 | 1.47 | 10 |
| Drawing-Black | 21.6 | 10 | 1.36 | 7 |
| Table-Black | 20.5 | 11 | 1.67 | 12 |
| Text-Black | 19.7 | 12 | 1.61 | 11 |

certainty surrounding the choice of "neutral," where the normal set was to respond in terms either of a positive or a negative vector, that is, the "neutral" response tended to represent a balance of positive and negative affect rather than a simple lack of affect. The paired-comparison procedure has the virtue that the judgments demanded of S are basically simpler (since the recall component is greatly reduced), and the repeated exposures to all of the stimuli tend to reduce any ambiguity relating to any given stimulus as well as to establish a relatively well-defined frame of reference in which to order the affective judgments. When the artifacts are thus removed, the result is the observed systematic, near-linear relationship between response latency and affective value.

From the applications point of view, it is conceivable that choice-response latencies might prove to be both more reliable and valid than the actual overt choices themselves, since the latencies are unaffected by inadvertent keyboard errors, and may be less prone to errors relating to Ss' expectations of what they think E may want them to prefer.

REFERENCES

- BERGUM, B., & LEHR, D. J. *Prediction of Stimulus Approach: Core Measures Experiment I*. (Research Report R66-8) Rochester, N. Y.: Xerox Corporation, 1966.
- BERGUM, B., & LEHR, D. J. Affect level, capillary

- pulse pressure, and response latency. *Journal of Applied Psychology*, 1967, **51**, 316-319.
- BERGUM, B., LEHR, D. J., & DOOLEY, R. P. *Prediction of stimulus approach: Core measures Experiment VI*. (Research Report R67-9) Rochester, N. Y.: Xerox Corporation, 1967.
- LEHR, D. J., BERGUM, B., & STANDING, T. E. Response latency as a function of stimulus affect and presentation order. *Perceptual and Motor Skills*, 1966, **23**, 1111-1116.

(Received August 23, 1968)

Manuscripts Accepted for Publication in the
Journal of Applied Psychology

- Psychodynamics of Intrinsic and Extrinsic Job Orientation: S. D. Saleh* and T. G. Grygier: Management and Systems Engineering, University of Waterloo, Waterloo, Ontario, Canada.
- Use of Leadership Powers in Industry: David Kipnis* and Joseph Cosentino: Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.
- A Situational Approach to Leadership Effectiveness: Walter Hill*: College of Business Administration, University of Florida, Gainesville, Florida 32601.
- Ability Correlates in Learning a Psychomotor Task: J. R. Hinrichs*: IBM Corporation, 112 East Post Road, White Plains, New York 10601.
- Some Intrapersonal Response Correlates of Organizational Rank: Cary M. Lichtman*: Department of Psychology, Wayne State University, Detroit, Michigan 48202.
- Effect of Training on the Productivity and Tenure of Sewing Machine Operators: Joel Lefkowitz*: 63 East 9th Street, New York, New York 10003.
- Leadership Beliefs of Australian Managers: Alfred W. Clark* and Sue McCabe: School of Applied Psychology, The University of New South Wales, P. O. Box 1, Kensington, New South Wales 2033.
- Comparison of Sources of Personal Satisfaction and of Work Motivation: Paul F. Wernimont*, Paul Toren, and Henry Kapell: Personnel Research Department, 3M Company, Building 224-1E, St. Paul, Minnesota 55101.
- Vocational Interests of Women: A Longitudinal Study of the SVIB: Earl Notling, Jr.,*: Counseling Center, University of Wisconsin, 415 West Gilman Street, Madison, Wisconsin 53706.
- A Comparative Study of Static versus Dynamic Scoring of Performance Tests for Industrial Operators: Gavriel Salvendy*, W. Douglas Seymour, E. Nigel Corlett: Department of Industrial Engineering, State University of New York at Buffalo, Buffalo, New York 14214.
- Satisfaction and Morale of Foremen as a Function of Middle Manager's Performance: Michael J. Kavanagh*, Arthur C. MacKinney, and Leroy Wolins: Department of Psychology, Iowa State University, Ames, Iowa 50010.
- Weighting Agreement Responses by Item Scale Values: Stephen Wollack, Jan P. Wijting, James G. Goodale, and Patricia Cain Smith*: Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43402.

* Asterisk indicates author for whom address is supplied.

A VALIDATION STUDY OF POLYGRAPH EXAMINER JUDGMENTS¹

PHILIP J. BERSH²

Temple University

The lie detection judgments of polygraph examiners in criminal investigations conducted by the military services were validated against unanimous guilt-innocence decisions by a panel of four Judge Advocate General (JAG) attorneys. Since the study did not permit isolation of the role played by the polygraph record itself, the examiner's judgment was considered the end product of his complete interrogation of a suspect. Each JAG attorney made an independent decision based upon perusal of case files from which all polygraph references were deleted. Attorneys were instructed to eliminate files lacking sufficient evidence and to disregard legal technicalities. Level of agreement was 92.4%. Percentage of agreement decreased significantly to 74.6% when the criterion was a majority JAG panel decision. The study supports the use by the military services of polygraph examiner judgments as an aid in determining whether to continue or to terminate the investigation of a suspect.

Criminal investigations conducted by the military services may include interrogations of suspects with the aid of the polygraph. In such cases the polygraph examiner's judgment concerning the truth of the suspect's replies to polygraph test questions often determines whether the investigation of the suspect should be continued or terminated. Yet a recent survey of lie detection by the polygraph method (Orlansky, 1964) has pointed to the almost complete lack of objective evidence bearing upon its reliability and validity for applications of this kind. Laboratory experiments, including those in which crimes have been simulated (e.g., Davidson,

1968; Kubis, 1962), have generally tended to support the effectiveness of the polygraph, and in particular that of the GSR indicator, for the detection of deception. However, it is often contended that the results of such experiments are inapplicable to live cases because of presumed radical differences in S's motivation to deceive and in his overall level of emotion.

One major advantage of laboratory over real-life studies of the polygraph is that laboratory controls make it possible to insure that lie detection judgments are based solely on the polygraph record. In such a context as a criminal investigation, on the other hand, the contribution made by the polygraph itself to the detection of deception is extremely difficult to isolate. The polygraph examination proper is embedded in an interrogation that includes a pretest interview and sometimes a posttest interrogation. In addition, the polygraph examiner has access to the case file and has ample opportunity to interact with the criminal investigator prior to the conduct of the examination. Each of these extrapolygraph forms of information is a potential source of cues which may have significant influence upon the examiner's judgment.

The present study was performed to assess the validity of lie detection judgments made by polygraph examiners in criminal investigations conducted by the military services. No

¹ The study was sponsored by the Department of Defense Research & Engineering Joint Working Group on Lie Detection, of which the author was a member. Data collection was carried out by Robert Brisentine, Office of the Provost Marshal General, United States Army, also a member of the Working Group.

Special thanks are due to S. Rains Wallace, original chairman of the Working Group, and to Jesse Orlansky, current acting chairman of the Group, for their valuable comments and suggestions throughout the course of this investigation. Helpful suggestions were also received from other members of the group.

² The study was performed while the author was employed by the United States Army Behavioral Science Research Laboratory, Washington, D. C. Requests for reprints should be sent to the author, Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.

TABLE 1

NUMBER OF CASES IN FINAL VALIDATION SAMPLE

| Zone of Comparison | | General Question Test | |
|---------------------|------------------------|-----------------------|------------------------|
| Deception indicated | No deception indicated | Deception indicated | No deception indicated |
| 37 | 52 | 35 | 33 |

attempt was made to disentangle the influence of the polygraph examination and record from that of the extrapolygraph sources of information available to the examiner. This would require a far more elaborate and expensive study than the one reported here. Accordingly, the data of the study bear only upon the validity of the examiner's judgement, not upon the validity of the polygraph method or of the polygraph record itself. In the final analysis it is this judgment, and not the record, which influences any further action that may stem from the interrogation. Validation of that judgment is required to determine whether its use for such purposes is warranted.

METHOD

Selection of Cases

Cases were drawn at random from a pool of criminal investigations conducted by the three branches of service during the years 1963-66. Selection of cases was subject to the following restrictions: (a) Cases judged "indeterminate" by the polygraph examiner were eliminated. (b) Half the cases represented the General Question Test (GQT) and the other half the Zone of Comparison (ZOC) type of polygraph examination.³ (c) Within each examination type, there was an equal number of Deception Indicated (DI) and of No Deception Indicated (NDI) judgments by the polygraph examiner.

Attrition due to the nature of the criterion used, as described below, affected the four resulting categories differentially. Table 1 presents the final number of cases in each category.

³The GQT type of examination begins with a control question but thereafter presents control and relevant questions in random order. In the ZOC type of examination, each relevant question is interpolated between a pair of control questions. The polygraph response to a relevant question is compared only with its surrounding control questions. In the case of the GQT type of examination, the polygraph response to a relevant question is compared with the level of response to control questions in general.

Validation Procedure

Selection of an appropriate criterion against which to validate the polygraph examiner's judgment poses considerable difficulty. Obviously, *prima facie* evidence of guilt or innocence of a suspect would be ideal. Unfortunately, such evidence is usually not available. Court-martial decisions constitute a reasonable possibility for criterion use. However, legal technicalities and sufficiency of evidence, factors unrelated to the question of whether a suspect actually committed a crime, often play a key role in these decisions. Their influence upon the criterion might artifactually reduce the validity of the examiner's judgment. Confessions also merit consideration as the criterion, but experience has demonstrated that some confessions are false. Equally important, they provide at best only a partial criterion, since their occurrence is interpreted as proof of guilt, but their failure to occur is not proof of innocence. Thus, lack of confession is neither confirming for NDI judgments, nor disconfirming for DI judgments. If cases selected for the validation sample are restricted to those in which a confession has occurred, then NDI judgments cannot be adequately validated. Such judgments are confirmed by proof or evidence of innocence; confessions only provide a basis for disconfirming them. The converse obviously holds for DI judgments.

In view of the deficiencies of the above criteria, the following validation procedure was adopted for the present study.

Polygraph records were removed from the case file, and all references to the polygraph were deleted. The case files were then submitted for review to a panel of four JAG attorneys. The use of a four-member panel was based on a preliminary study involving nine JAG attorneys representing all three service branches. This study demonstrated that unanimity among four attorneys meant unanimity among all nine attorneys. On the other hand, unanimity for a three-member panel did not assure unanimity for the full nine-member panel. Each member of a panel reviewed and judged the cases independently. The attorneys were given explicit instructions to disregard all legal technicalities and to judge each case solely on the evidence contained in the file. As a precaution, each attorney was first required to eliminate files containing, in his judgment, insufficient evidence to warrant a positive determination of guilt or innocence. Otherwise, cases judged DI by the polygraph examiner might be judged not guilty by the JAG panel merely because the case files contained little information. Only those cases which resulted in a unanimous decision by the JAG panel were retained in the validation sample.

Thus, the validity of the polygraph examiner's judgments was estimated by determining the level of their agreement with the unanimous decisions of a panel of four JAG attorneys, each of whom made an independent judgment of the guilt or innocence of suspects. Unanimous agreement among legal experts experienced at sifting evidence and instructed

TABLE 2

COMPARISON OF POLYGRAPH EXAMINER AND
UNANIMOUS JAG PANEL DECISIONS
IN GQT CASES

| JAG panel | Polygraph examiner | | |
|------------|---------------------|------------------------|-------|
| | Deception indicated | No deception indicated | Total |
| Guilty | 31 | 1 | 32 |
| Not guilty | 4 | 32 | 36 |
| Total | 35 | 33 | 68 |

TABLE 3

COMPARISON OF POLYGRAPH EXAMINER AND
UNANIMOUS JAG PANEL DECISIONS
IN ZOC CASES

| JAG panel | Polygraph examiner | | |
|------------|---------------------|------------------------|-------|
| | Deception indicated | No deception indicated | Total |
| Guilty | 34 | 4 | 38 |
| Not guilty | 3 | 48 | 51 |
| Total | 37 | 52 | 89 |

both to disregard technicalities and to eliminate cases where sufficient information is lacking would seem to provide a criterion which approaches the ideal of direct proof of guilt or innocence.

An initial group of 227 case files was submitted to a panel of four United States Army JAG attorneys. When attrition reduced the number of usable cases below required levels, an additional group of 96 case files was submitted to a panel of four United States Air Force JAG attorneys. Seventy-eight of the first 227 and 2 of the later 96 cases were eliminated by panel members for lack of sufficient information in the case files.⁴ The panel's decision was unanimous in 91 of the 149 cases remaining from the first set and in 66 of the 94 cases remaining from the second set.

RESULTS

Tables 2 and 3 summarize the data for cases involving the GQT and ZOC types of examination, respectively. In Table 4 corresponding cell entries for Tables 2 and 3 have been combined.

The percentages of agreement between polygraph examiner and JAG panel are 92.6 for GQT cases, 91.0 for ZOC cases, and 92.4 for all cases combined. Agreement of 90.3% was reached on cases judged DI by the polygraph examiner and 94.1% on cases judged NDI. Chi-square tests of independence were performed on the data in the three tables. Chi-square for Table 2 is 49.7, for Table 3 62.6, and for Table 4 112.4. These chi-squares are all significant at well beyond the .001 level. Chi-square ($df = 1$, $p < .001$) is

⁴ The discrepancy between the Army and Air Force JAG Panel rejection rate was to a considerable extent due to a single member of the Army Panel. It would have been less justifiable to eliminate this panel member post facto than to tolerate his unusually stringent criterion for sufficiency of evidence.

10.8. Phi coefficients were also computed for the three tables. For the GQT cases, $\Phi = .86$ ($\Phi_{\max} = .97$); for the ZOC cases, $\Phi = .84$ ($\Phi_{\max} = .97$); and for all cases combined, $\Phi = .85$ ($\Phi_{\max} = .97$).

Unanimity on the part of the JAG panel is, of course, a stringent criterion. Generally, the initial decision of conviction by a court-martial is the product of a two-thirds majority of the court members, and, upon subsequent review, decision is by a simple majority of the appellate tribunal. It seems worthwhile, therefore, to consider also data for cases in which the JAG panel decision was a majority one. (This, of course, is equivalent to a decision by three-fourths of the panel.) There were 59 such cases.

In view of the small number of cases involved, only the combined data for all cases are presented in Table 5.

The percentage of agreement between polygraph examiner and a majority of the JAG panel is 74.6%. Chi-square for Table 5 is 14.7 ($p < .001$), and the corresponding phi coefficient is .49 ($\Phi_{\max} = .87$).

TABLE 4

COMPARISON OF POLYGRAPH EXAMINER AND
UNANIMOUS JAG PANEL DECISIONS
IN ALL CASES

| JAG panel | Polygraph examiner | | |
|------------|---------------------|------------------------|-------|
| | Deception indicated | No deception indicated | Total |
| Guilty | 65 | 5 | 70 |
| Not guilty | 7 | 80 | 87 |
| Total | 72 | 85 | 157 |

TABLE 5

COMPARISON OF POLYGRAPH EXAMINER AND
MAJORITY JAG PANEL DECISIONS
IN ALL CASES

| JAG panel | Polygraph examiner | | |
|------------|---------------------|------------------------|-------|
| | Deception indicated | No deception indicated | Total |
| Guilty | 24 | 10 | 34 |
| Not guilty | 5 | 20 | 25 |
| Total | 29 | 30 | 59 |

Table 6 compares levels of agreement for cases in which the JAG panel decision was unanimous with those in which its decision was a majority one.

Finally, the majority decision cases were combined with the unanimous decision cases to provide level of agreement data for all cases involving at least a majority decision by the JAG panel. The data for all cases are presented in Table 7.

The percentage of agreement is 87.5%, chi-square for Table 7 is 121.6 ($p < .001$), and the corresponding phi coefficient is .75 ($\Phi_{\max} = .97$). Cases judged DI by the polygraph examiner yielded 88.1% agreement and those judged NDI resulted in 86.9% agreement.

DISCUSSION

The data show clearly that the polygraph examiner's judgment is predictive of the JAG panel decision. This is particularly true in cases where the evidence is sufficiently compelling to produce a unanimous decision by the panel. Level of agreement between the examiner and the panel decreases significantly for cases in which the panel's decision is

TABLE 6

COMPARISON OF LEVELS OF AGREEMENT
BETWEEN POLYGRAPH EXAMINER
AND JAG PANEL

| JAG panel | Agree | Disagree | Total |
|---------------------|-------|----------|-------|
| Unanimous decisions | 145 | 12 | 157 |
| Majority decisions | 44 | 15 | 59 |
| Total | 189 | 27 | 216 |

Note.— $\chi^2 = 12.3, p < .001$.

TABLE 7

COMPARISON OF POLYGRAPH EXAMINER AND
MAJORITY OR UNANIMOUS JAG PANEL
DECISIONS IN ALL CASES

| JAG panel | Polygraph examiner | | |
|------------|---------------------|------------------------|-------|
| | Deception indicated | No deception indicated | Total |
| Guilty | 89 | 15 | 104 |
| Not guilty | 12 | 100 | 112 |
| Total | 101 | 115 | 216 |

simply a majority one. A parsimonious interpretation of these relationships between examiner and panel judgments would appear to implicate the case file as the only common source of information. In unanimous decision cases the file evidence may be more convincing for the examiner, as it is by definition for the panel, than in majority decision cases. It is very unlikely, however, that the examiner's judgment is determined solely by the file, or even by the file in combination with information provided by the criminal investigator. The polygraph examination is ordinarily not given to suspects whose guilt or innocence has already been substantially or finally established. Only where real doubt exists about the guilt status of the suspect is he permitted or asked to volunteer for an examination. The fact is that the case file at the time of the polygraph examination was less complete, and often far less complete, than when it was delivered to the JAG panel. The examiner's judgment, then, presumably reflects also the influence of other sources of information, such as the pretest interview, the polygraph examination proper, and the polygraph charts. As noted earlier, the design of the present study makes it impossible to determine the relative contribution to the examiner's judgment and to the validity of that judgment made by each of the possible sources of information available to him.

In particular, no conclusions are drawn about the validity of the polygraph record itself or about its contribution to the validity of the examiner's judgment. It seems reasonable to assume, however, that the validity of that judgment sets an upper limit for the

validity of the record. Since agreement of the examiner's judgment with the criterion (i.e., unanimous JAG panel decision) is high, a substantial validity for the polygraph record is certainly possible. A study which investigates the joint and independent influence of the information sources available to the examiner (with emphasis on the role of the polygraph charts) has been planned, but is awaiting approval by the Department of Defense, as well as the removal of practical obstacles to efficient data collection.

Caution dictates that any conclusions drawn should be limited to criminal investigations carried out by the military services. They should not be applied to personnel screening by the services or by other governmental agencies. Nor should they be generalized to use of the polygraph by individuals or agencies outside the federal government. The latter restriction is especially justified by the marked differences in favor of the military services with respect to quality control over the polygraph examiner's training and performance. Of course, wherever civilian agencies employ standards approaching those of the military services, the findings of this study will presumably apply.

A final comment is in order about the nature of the validation carried out in this

study. As far as the polygraph examiner is concerned, he has made judgments about the truth of the suspect's responses. On the other hand, the criterion itself is concerned with guilt or innocence determination. Accordingly, the labels "DI" and "NDI" by which the examiner categorizes suspects appear to be reasonably valid indicators of the guilt or innocence of the suspect, particularly in cases where the criterion decision is unanimous.

As noted in the introduction, the polygraph interrogation is used within the military services to determine whether to continue or terminate the investigation of particular suspects in criminal cases where the evidence on hand is inconclusive. Whatever the basis may be for the examiner's judgment, its use for such a purpose is strongly supported by the present study.

REFERENCES

- DAVIDSON, P. O. Validity of the guilty-knowledge technique: The effects of motivation. *Journal of Applied Psychology*, 1968, **52**, 61-65.
- KUBIS, J. F. *Studies in lie detection: Computer feasibility considerations*. (Tech. Rep. No. 62-205) Rome Air Development Center, 1962.
- ORLANSKY, J. *An assessment of lie detection capability*. (Tech. Rep. No. 62-16, declassified version) Washington, D. C. Institute of Defense Analyses, Research and Engineering Support Division, 1964.

(Received August 28, 1968)

MMPI SCORES AS RELATED TO AGE, EDUCATION, AND INTELLIGENCE AMONG MALE JOB APPLICANTS

FRED J. THUMIN¹

University of Missouri at St. Louis

This study was designed to ascertain the relation of age, education, and intelligence to the 13 basic scales of the MMPI among normal male Ss in a competitive employment situation. Correlational analysis revealed that (a) age was negatively related to Scales *F*, 7, and 8, (b) with age and intelligence partialled out, education was positively related to Scales *L*, *K*, and 9, but negatively to Scale 0, and (c) with age and education partialled out, intelligence was negatively related to *L* and positively to 5. All multiple correlations using age, education, and intelligence as predictor variables and MMPI scales as criterion variables resulted in significant correlations except those involving Scales 1 and 2.

The relationship between MMPI responses and the age, education, and intelligence of Ss has been studied by a number of investigators, but as yet the data are too limited to permit either cogent conclusions or firm generalizations regarding the interactions of these variables. While certain of the findings are reasonably consistent from one report to another, others are seemingly inconsistent or, in some cases, contradictory. At least in part, the discrepancies may be attributed to differences in the nature of Ss, but failure to control or to partial out the influence of confounding variables is undoubtedly an additional contributing factor.

Nonetheless, among relatively diverse populations, the age variable was found to be positively related to Scales 1 and 2 (Aaronson, 1958; Brozek, 1955; Calden & Hokanson, 1959), to Scale 0 (Brozek, 1955; Calden et al., 1959), and to Scale 5 among male Ss (Applezweig, 1953; Brozek, 1955). Similarly, negative relationships between age and Scales 8 and 9 have been reported by Applezweig (1953), Brozek (1955), and Gynther and Shimkunas (1966).

Regarding education, perhaps the most consistent finding is that among male Ss, as educational level increases, a corresponding elevation appears on Scale 5 (Applezweig,

1953; Brehm, 1954; Gough, 1954; Gynther et al., 1966). The relation between intelligence and MMPI performance is anything but clear, although Stanton (1956), Gynther et al. (1966), and others have found that, with male Ss, higher levels of intelligence are accompanied by higher scores on Scale 5. Although the foregoing relationships are among those most frequently and consistently reported, it should be pointed out that exceptions could be cited in every case, revealing the still very amorphous nature of the situation.

The objective of the present study was to ascertain, among normal individuals in a competitive employment setting, the relationship between the basic scales of the MMPI and the age, education, and intelligence of Ss. Unlike most earlier studies in this area, a special attempt was made to separate or isolate the effects of the three personal variables under consideration.

METHOD

The sample consisted of 236 male Caucasians between the ages of 19 and 56 who had been referred to a psychological consulting firm by 36 different business organizations for testing and interviewing between September, 1966, and June, 1968. Although the great majority of the men were applying for positions with the firms which referred them for evaluation, some were current employees being considered for promotion or advancement. The Ss were evaluated for a variety of positions, including managerial, administrative, accounting, engineering, sales, and foreman. From an experimental standpoint,

¹Requests for reprints should be sent to the author, School of Business Administration, University of Missouri, 8001 Natural Bridge Road, St. Louis, Missouri 63121.

TABLE 1
MMPI SCALES AS RELATED TO AGE, EDUCATION, AND INTELLIGENCE

| Scale | Column A— Correlation coefficient | | | Column B— Partial r | | | Column C— Multiple R | M | SD |
|--------|--------------------------------------|-----------|--------------|--------------------------|-----------|--------------|---------------------------|------|------|
| | Age | Education | Intelligence | Age | Education | Intelligence | | | |
| L | .02 | .09 | -.15* | .0 | .19** | -.22** | .24** | 4.3 | 2.4 |
| F | -.18** | -.04 | .09 | -.17* | -.11 | .10 | .22** | 2.3 | 1.9 |
| K | -.08 | .20** | .11 | -.04 | .17* | .01 | .21** | 17.9 | 4.6 |
| 1 (Hs) | -.08 | .02 | .0 | -.08 | .02 | -.02 | .08 | 11.5 | 2.7 |
| 2 (D) | .01 | -.12 | -.10 | -.02 | -.09 | -.04 | .13 | 16.6 | 3.1 |
| 3 (Hy) | -.01 | .14* | .13* | .02 | .09 | .07 | .16* | 20.4 | 3.4 |
| 4 (Pd) | -.08 | .01 | .12 | -.06 | -.06 | .12 | .14* | 22.0 | 3.7 |
| 5 (Mf) | .08 | .12 | .19** | .13 | .05 | .17* | .23** | 23.0 | 4.0 |
| 6 (Pa) | -.02 | .18** | .17** | .03 | .12 | .10 | .21** | 8.8 | 2.1 |
| 7 (Pt) | -.18** | .03 | .08 | -.17* | -.03 | .05 | .18** | 24.0 | 3.3 |
| 8 (Sc) | -.20** | .10 | .12 | -.18** | .03 | .05 | .22** | 23.1 | 4.2 |
| 9 (Ma) | -.12 | .22** | .16* | -.08 | .16* | .05 | .24** | 19.8 | 3.8 |
| 0 (Si) | .09 | -.26** | -.11 | .05 | -.24** | .03 | .27** | 17.4 | 6.8 |

* $p < .05$.** $p < .01$.

It is significant that Ss were similar and homogeneous from the standpoint of motivation and commonness of objective; that is, all were tested in the same realistic employment situation, and presumably all were strongly motivated to perform as well as possible on the tests.

As part of the evaluation process, all Ss were administered the Otis Self-Administering Test of Mental Ability (Higher Examination, Form A) and the MMPI. Only the 13 basic scales of the MMPI were used for purposes of the present investigation. The Ss were also required to complete an employee inventory from which information pertaining to their chronological age and years of formal education was obtained. The mean age, education, and Otis raw score for the sample were 34.5 yr., 13.9 yr., and 52.6 points, respectively, with corresponding SD s of 7.9, 2.5, and 10.9. An Otis raw score of 52.6 converts to a percentile score of approximately 82, using general adult population norms.

The data were analyzed first by obtaining simple product-moment correlations between the MMPI scores and the age, education, and intelligence of Ss. In addition, a multiple regression analysis was performed for each MMPI scale using age, education, and intelligence as predictor variables. This analysis resulted in partial correlations which revealed the separate and independent effects of the predictor variables upon the scales, as well as multiple correlations which provided information as to the combined effects of the predictors.

RESULTS AND DISCUSSION

Table 1 shows the mean K-corrected raw score for each MMPI scale, the correspond-

ing SD s, and the correlation coefficients between the scales and the three personal variables (age, education, and intelligence). The simple correlation coefficients appear in Column A, the partial coefficients in Column B, and the multiple coefficients (using age, education, and intelligence as predictor variables) in Column C. Although the data do not appear in Table 1, age correlated $-.19$ with education, and $-.21$ with intelligence, while the correlation between education and intelligence was $.48$. All three values are significant at or beyond the $.01$ level of confidence. The negative relationship between age and education can probably be explained in terms of the greater inclination for younger individuals to carry their formal education further than did the members of the previous generation.

It is noted in Column A that age correlated significantly with 3 of the 13 basic scales (viz., with F , 7, and 8), and all were in the negative direction. Moreover, these relationships were virtually identical in Columns A and B, indicating that it made little difference whether or not the variables of education and intelligence were partialled out. The inverse relations between age and Scales F and 8 are consistent with the recent findings of

Gynther et al. (1966) and may well reflect an individual's inclination to become increasingly practical and realistic in his thinking as he grows older, at least through middle age.

Perhaps even more interesting, however, is the fact that the significant relations between age and Scales 1, 2, 5, 9, and 0, reported by several other investigators, did not occur. There is no suggestion in the present data that individuals between early adulthood and middle age become increasingly concerned with bodily ailments, increasingly despondent, lose zest and enthusiasm, or grow socially introverted. On the other hand, the present sample contained relatively few Ss beyond middle age, so that many of the traits typically associated with advancing age simply may have gone undetected. Also, it can not be overlooked that this sample consisted of job applicants and employees being considered for promotion, as opposed to college students and hospital populations as were used in most previous studies. And there is evidence that elevations on a given scale may have different meanings or implications for different groups (Thumin, 1965). For example, a high Mf score among advertising men may reflect artistic and aesthetic interests rather than homosexual tendencies, as might be the case with certain other groups of Ss.

It is perhaps not surprising to find that, as education increased, there was a corresponding decrement on Scale 0, suggesting that formal education facilitates or enhances the capacity and/or desire for social interaction. It seems equally feasible that a positive relation existed between education and Scale 9 for those individuals with substantial amounts of enthusiasm, optimism, and energy might well be expected to carry their formal education further than would persons with somewhat fewer endowments along these lines.

In several instances, the interaction effects of education and intelligence are apparent. For example, in Column A both Hy and Pa are significantly related to education and intelligence, whereas the "uncontaminated" values in Column B drop below the level re-

quired for significance. Moreover, the *L* scale is negatively correlated with intelligence in both Columns A and B, but positively (and significantly) correlated with education only in Column B where the influence of intelligence was partialled out. Apparently the brighter people with more education are able to "see through" and avoid the unrealistically favorable items comprising the *L* scale, whereas better educated people, minus the insight afforded by higher intelligence, are considerably more apt to subscribe to these kinds of items in their efforts to present themselves in the most possible favorable light to a prospective employer. Although the present study failed to reveal a significant relation between education and Scale 5 (as reported by a number of other investigators), the expected positive correlation between intelligence and Scale 5 did occur. As a matter of fact, with age and education partialled out, intelligence correlated significantly with only two scales—5 and *L*.

Finally, it is noteworthy that 11 of the 13 multiple correlations were statistically significant, and most of them at or beyond the .01 level. Thus, when age, education, and intelligence were used in combination as predictor variables and weighted so as to maximize their relation to the criterion variables (i.e., to the MMPI scales), nearly all were found to be of predictive value. The three highest correlations were with Scales *L* (.24), 9 (.24), and 0 (.27).

REFERENCES

- AARONSON, B. S. Age and sex influences on MMPI profile peak distributions in an abnormal population. *Journal of Consulting Psychology*, 1958, 22, 203-206.
- APPLEZWEIG, M. H. Educational levels and Minnesota Multiphasic profiles. *Journal of Clinical Psychology*, 1953, 9, 340-344.
- BREHM, M. L. An examination of MMPI results and delinquency data for college and noncollege groups. Unpublished manuscript, 1954. Cited by W. G. Dahlstrom and G. S. Welsh, *An MMPI Handbook*, Minneapolis, The University of Minnesota Press, 1960, p. 312.
- BROZEK, J. Personality changes with age: An item analysis of the MMPI. *Journal of Gerontology*, 1955, 10, 194-206.

- CALDEN, G., & HOKANSON, J. E. The influence of age on MMPI responses. *Journal of Clinical Psychology*, 1959, 15, 194-195.
- GOUGH, H. G. Some personality differences between high ability high school students who do, and do not, go to college, *American Psychologist*. 1954, 9, 559.
- GYNTHIER, M. D., & SHIMKUNAS, A. M. Age and MMPI performance. *Journal of Consulting Psychology*, 1966, 30, 118-121.
- STANTON, J. M. Group personality profile related to aspects of antisocial behavior. *Journal of Criminal Law, Criminology & Police Science*, 1956, 47, 340-349.
- THUMIN, F. J. Personality characteristics of diverse occupational groups. *Personnel and Guidance Journal*, 1965, XLIII, 5, 468-470.

(Received September 6, 1968)

THE SVIB FOR WOMEN AND DEMOGRAPHIC VARIABLES IN THE PREDICTION OF OCCUPATIONAL TENURE

THOMAS H. STONE AND GARY T. ATHELSTAN¹

Center for Research and Education in the Health Occupations

The American Rehabilitation Foundation

In an ex post facto attempt to predict women's "occupational tenure," 198 occupational therapists and 255 physical therapists were studied. Eight demographic variables and selected scales of the women's Strong Vocational Interest Blank (SVIB) were incorporated in two types of prediction equations. A double cross-validation design was used to develop and test four multiple regression and four reciprocal averages equations. Five of the demographic variables correlated significantly with tenure, but none of the SVIB scales proved to be stable predictors. The method of reciprocal averages prediction yielded equations which proved more stable across samples and suffered less shrinkage than those produced by the multiple regression technique.

The prediction of "occupational tenure" for women, defined here as the percentage of time worked in a field after completion of training for that field, has become an important manpower research problem in recent years. A reliable method of predicting which women will have the greatest occupational tenure over their working lives, and who will therefore provide the greatest "return on educational investment," would have considerable practical value for educators. While the SVIB for Women has not been used to predict tenure directly, several studies have related SVIB scores to tenure-related criteria.

A follow-up study of two groups of high scorers on the SVIB Social Worker and Laboratory Technician scales (Harmon, 1968) found that the scales predicted "usual occupation" reasonably well for those Ss who claimed one; however, the SVIB was not successful in predicting whether or not women would report a "usual career" 10 to 14 yr. after entrance to college. Another study (Harmon, 1967) utilized a behavioral criterion of tenure, hypothesizing a relationship between women's working patterns and their SVIB Housewife and Own occupational scale

scores. No differences were found on the Housewife scale between groups having different work patterns, and only one group differed in the expected direction from others on Own scale. Precollege SVIB scores were found to have some success in predicting college major and occupation in a seven-year follow-up study by Nolting (1967) of 316 University of Minnesota female graduates. Despite the lack of comparability of tenure indexes among these three studies, it appears that the SVIB is, at best, only moderately related to occupational tenure for women.

Several important health occupations whose members are mostly women, including occupational and physical therapy, are now suffering from critical personnel shortages. Since only about one-half of the women trained in these fields are working in them (Flint, 1968), it appears that retention of trained personnel may do more to help alleviate the manpower shortage than recruitment or any of the other traditional solutions. Therefore, health manpower investigators are attempting to identify factors which are predictive of tenure, and which can be used to select potential high-tenure people for training and employment. This study compares the relative efficiency of selected SVIB scales and demographic variables in the ex post facto prediction of tenure.

¹ Requests for reprints should be sent to Gary T. Athelstan, Assistant Director of Research, Center for Research and Education in the Health Occupations, The American Rehabilitation Foundation, 1800 Chicago Avenue, Minneapolis, Minnesota 55404.

METHOD

Subjects

Mailing lists were compiled from the membership rosters of professional organizations for occupational therapists (OTs) and physical therapists (PTs), and from state certification records dating back to 1960. In 1966, each *S* identified in this manner was mailed a Career Patterns Questionnaire and the women's SVIB (Form TW400-R, 1966) as part of a survey of health manpower in Minnesota (Flint, 1968). Three follow-ups were made over a period of about 6 wk., yielding returns from 198 OTs and 255 PTs. The respondents constitute nearly 80% of the qualified (fully trained) females in these fields in the state of Minnesota.

Tenure Index

An index of tenure was constructed from the questionnaire data available on each *S*. The index was operationally defined as the total number of years of professional work experience divided by the total number of years since graduation from the professional program. The values obtained were then converted to percentages so that the scale was continuous from 0 to 100%. This index permits women of different ages to be compared on the same measure, and is a convenient way of representing "return on educational investment."

Demographic Predictors

Eight demographic variables were selected from the questionnaires as likely predictors of tenure for both OT and PT groups. These were marital status; year of birth; amount of own income; amount of spouse's income; when the decision to become an OT or PT was made, ranging from "before entering high school" to "after completing college"; number of children living at home; presence of children in the 0-5 age group; and presence of children in the 6-12 age group.

Selected SVIB Scales

A preliminary step-wise regression analysis of the relationship between tenure and all scales of the women's SVIB was performed to select the scales most likely to predict the criterion. Four SVIB scales yielded significant correlations ($p < .05$) with tenure for the PTs: the Lawyer, Music Teacher, and Speech Pathologist scales, and Experimental Scale No. 2.² Only one predictor emerged for the OT group: SVIB Experimental Scale No. 2.

In order to increase the number of scales to be used in the analysis, the Occupational Therapist and College Physical Education Teacher scales were added to the possible predictors for the OT group on a purely intuitive basis. The Math-Science Teacher scale was similarly included for PTs, and the Housewife and Social Worker scales, which were used by

Harmon (1967, 1968), were selected as possible predictors for both groups.

Procedure

Two methods of prediction were used in the study: linear multiple regression and reciprocal averages prediction (Weiss, 1963). The investigators were also interested in comparing the relative effectiveness of the two prediction techniques. It was felt that reciprocal averages prediction (RAP) might be a better method of prediction since it does not assume linearity and categorical data may be used. Also, some evidence (Weiss & Dawis, 1968) exists which suggests that RAP developmental prediction equations are more stable in cross-validation than are multiple regression equations.

A double cross-validation design (Mosier, 1951) was used in which the OT and PT groups were each randomly split into two approximately equal-sized groups. In this design, eight prediction equations were developed, one multiple regression and one RAP equation for each of the four development groups. Each of the equations incorporated all of the selected demographic and SVIB variables. Within the OT and PT categories, each development group served as the other's cross-validation group.

RESULTS

Forty-nine percent of the OTs and 69% of the PTs were employed within their respective fields at the time of the survey. The OTs had worked as OTs an average of 59% of the time since completing their training. The average occupational tenure reported by PTs was 68%.

Zero-order correlations between the various predictors and tenure for the OT and PT groups are reported in Tables 1 and 2. Inspection of both tables shows five demographic variables to be the best predictors of tenure when used singly: number of children; presence of children in the 0-5 age group; presence of children in the 6-12 age group; own income; and spouse's income. Number of children, which showed the highest correlation with the criterion, was found to be inversely related to tenure, as was presence of children in both the 0-5 and 6-12 age groups.

Tenure was positively related to own income and inversely related to spouse's income. Marital status correlated highly with tenure in both PT groups, but in neither of the OT groups. Among PTs, marriage was associated with low tenure. A significant correlation was found between age and the criterion for all groups, although the relationship for PTs was

²This scale is for research purposes and is not available for general use.

TABLE 1
ZERO-ORDER CORRELATIONS BETWEEN PREDICTORS
AND TENURE FOR OCCUPATIONAL THERAPISTS
(Total N = 198)

| Variable | % | <i>r</i> (Group 1) | <i>r</i> (Group 2) |
|-------------------------|----------|-----------------------|-----------------------|
| Marital status | | -.15 | -.19 |
| Never married | 20.2 | | |
| Married | 79.8 | | |
| Number of children | | -.53 | -.62 |
| None | 34.3 | | |
| One or two | 37.4 | | |
| Three or more | 28.3 | | |
| Own income | | .60 | .39 |
| Less than \$5000/year | 44.5 | | |
| \$5000 or more/year | 55.5 | | |
| Spouse's income | | -.33 | -.58 |
| None (includes un- | | | |
| married) | 27.8 | | |
| Less than \$10,000/ | | | |
| year | 44.9 | | |
| \$10,000 or more/year | 27.3 | | |
| When decide on OT | | .01 | -.17 |
| career | | | |
| High school or before | 36.5 | | |
| During college | 55.0 | | |
| After college | 8.5 | | |
| Have children aged 0-5 | 46.0 | -.42 | -.41 |
| Have children aged 6-12 | 66.7 | -.51 | -.50 |
| Variable | <i>M</i> | <i>r</i> (Group 1) | <i>r</i> (Group 2) |
| Age | 33.07 | -.28 | -.46 |
| SVIB OT | 46.14 | .00 | .08 |
| SVIB Housewife | 34.47 | -.18 | -.06 |
| SVIB College Phy. Ed. | | | |
| Teacher | 25.15 | .04 | .16 |
| SVIB Social Worker | 34.71 | .12 | -.07 |
| SVIB Experimental | | | |
| Scale No. 2 | 45.31 | .22 | .07 |
| Percentage of time | | | |
| worked | 58.84 | | |

curvilinear, and was not initially reflected in the correlation coefficient. Among the OTs age was negatively related to tenure: Younger women had a higher level of tenure than did older women. To some extent, the higher tenure of the younger women is an artifact of the index; it is easier for the young women to demonstrate high tenure, simply because they have had less time to participate in home-making and other activities which interfere with occupational tenure.

The only one of these demographic variables not having a significant zero-order cor-

relation with either OT or PT tenure was "When decide on OT-PT career?" None of the SVIB scales showed significant correlations with the criterion.

The RAP and multiple regression analyses produced results similar to the zero-order correlations. Predictors which appeared in only one of the two development groups within either of the OT or PT categories may be regarded as unstable. With the exception

TABLE 2
ZERO-ORDER CORRELATIONS BETWEEN PREDICTORS
AND TENURE FOR PHYSICAL THERAPISTS
(Total N = 255)

| Variable | % | <i>r</i> (Group 1) | <i>r</i> (Group 2) |
|-------------------------|----------|-----------------------|-----------------------|
| Marital status | | -.47 | -.40 |
| Never married | 33.7 | | |
| Married | 66.3 | | |
| Number of children | | -.65 | -.68 |
| None | 52.5 | | |
| One or two | 31.4 | | |
| Three or more | 16.1 | | |
| Own income | | .51 | .33 |
| Less than \$5000/year | 34.9 | | |
| \$5000 or more/year | 65.1 | | |
| Spouse's income | | -.54 | -.50 |
| None (includes un- | | | |
| married) | 38.4 | | |
| Less than \$10,000/ | | | |
| year | 39.6 | | |
| \$10,000 or more/year | 22.0 | | |
| When decide on PT | | -.06 | .01 |
| career | | | |
| High school or before | 31.0 | | |
| During college | 43.5 | | |
| After college | 25.5 | | |
| Have children aged 0-5 | 65.5 | -.52 | -.38 |
| Have children aged 6-12 | 77.6 | -.46 | -.60 |
| Variable | <i>M</i> | <i>r</i> (Group 1) | <i>r</i> (Group 2) |
| Age | 35.04 | -.03 | -.16 |
| SVIB Housewife | 34.44 | -.15 | -.13 |
| SVIB Social Worker | 31.39 | -.25 | -.15 |
| SVIB Experimental | | | |
| Scale No. 2 | 31.39 | .16 | .12 |
| SVIB Math-Science | | | |
| Teacher | 33.16 | .00 | .18 |
| SVIB Lawyer | 23.27 | -.03 | -.18 |
| SVIB Music Teacher | 23.04 | -.04 | -.08 |
| SVIB Speech Patholo- | | | |
| gist | 28.97 | .07 | -.15 |
| Percentage of time | | | |
| worked | 67.84 | | |

of "When decide on OT-PT career," all demographic variables were stable and appeared as such in both development groups of the OT and PT categories. None of the SVIB scales proved to be stable predictors. Several of the SVIB scales emerged as significant in one or another of the developmental equations, but all washed out upon cross-validation.

Two of the stable predictors, number of children and age, demonstrated a significant ($p < .05$) curvilinear relationship with the criterion. Number of children is nonlinear for all groups except OT Group 2. Increasing age is associated with decreasing tenure for PTs to age 37, at which point tenure begins to increase. This change probably reflects reentry to the labor market among women whose youngest child has entered school; however, the mean criterion score for the oldest PTs is not as high as that of the youngest PTs.

Table 3 illustrates the level of prediction attainable when the RAP and multiple regression techniques are used to combine optimally all SVIB scales and demographic variables. While the developmental multiple R 's are consistently higher than the reciprocal averages correlation coefficients, the reverse is true of the cross-validated coefficients. This result is consistent with the findings of other studies, and suggests that RAP equations are more stable across samples and suffer less shrinkage.

DISCUSSION

The results of this study suggest that demographic variables are better predictors of occupational tenure for these groups than are selected scales of the women's SVIB. A preliminary step-wise regression analysis suggested that several SVIB scales were significantly related to occupational tenure. However, none of the zero-order correlations between the SVIB scales and tenure were significant, and the scales added very little to the multiple regression or reciprocal averages prediction equations.

The data also encourage the use of reciprocal averages prediction in place of the usual linear multiple regression technique, especially when the nature of the predictor-criterion relationship is unknown and categorical pre-

TABLE 3
DEVELOPMENT AND CROSS-VALIDATED CORRELATIONS
BETWEEN ALL PREDICTORS AND TENURE, USING
MULTIPLE REGRESSION AND RECIPROCAL
AVERAGES PREDICTION

| Group | Multiple regression | | Reciprocal averages prediction | |
|-------|---------------------|----------------------|--------------------------------|----------------------|
| | Develop- ment | Cross- validation | Develop- ment | Cross- validation |
| OT 1 | .76 | .51 | .69 | .61 |
| OT 2 | .74 | .60 | .70 | .65 |
| PT 1 | .76 | .60 | .73 | .68 |
| PT 2 | .74 | .60 | .72 | .67 |

dictors are used. The RAP equations developed in this study identified and incorporated curvilinear relationships and also proved more stable than the multiple regression equations.

Although the demographic variables used in this study were found to be much better predictors of tenure than selected SVIB scales, it should be noted that they were used as ex post facto predictors: They related to tenure differences, but were not accessible in this study for longitudinal prediction. The one demographic factor that might have been useful in prediction, the time when occupational choice was made, turned out to be non-significant. On the other hand, the failure of the SVIB scales to relate to occupational tenure provides further evidence that the vocational behavior of women cannot be adequately accounted for by means which are usually successful for men.

One of the major difficulties in predicting the working patterns of women is accounting for the variance that results from their entering and leaving the labor market. It is likely that this behavior will seldom be reflected in scores on any of the standard occupational scales of the SVIB, since women leave an occupation more often for family and related reasons than for lack of interest in their jobs.

However, the negative results obtained to date in using the women's SVIB to predict occupational tenure should not discourage the use of this instrument in such studies. Some of the factors which differentiate women having high occupational tenure from those

having low tenure are almost certainly tapped by the SVIB, although not by the existing occupational scales. A very recent study (Schissel, 1968) holds promise for scales developed specifically to predict "career orientation." The most useful of several scales developed by Schissel differentiated women who had been employed at least five consecutive years from these who were not employed, with 41% distributional overlap between the groups.

Unfortunately, as Schissel's Career Orientation Scale for women utilizes the men's SVIB, it was not possible to evaluate the scale for the groups of OTs and PTs studied here. However, it is likely that a scale developed for the women's SVIB will work at least as well as Schissel's, possibly even better.

REFERENCES

FLINT, R. T., & SPENSLEY, K. C. Career patterns of Minnesota occupational therapists. *American Journal of Occupational Therapy*, 1968, 22, 30-34.

HARMON, L. W. Women's working patterns related to their SVIB Housewife and 'Own' Occupational scores. *Journal of Counseling Psychology*, 1967, 14, 299-301.

HARMON, L. W. Predictive power over ten years of measured social service and scientific interests among college women. *Journal of Applied Psychology*, 1969, 53, 193-198.

MOSTIER, C. I. Problems and designs for cross-validation. *Educational and Psychological Measurement*, 1951, 11, 5-11.

NOLTING, E., JR. A study of female vocational interests: Precollege to post-graduation. Unpublished doctoral dissertation, University of Minnesota, 1967.

SCHISSEL, R. F. Development of a career-orientation scale for women. *Journal of Counseling Psychology*, 1968, 15, 257-62.

WEISS, D. J. A technique for curvilinear multivariate prediction. Unpublished doctoral dissertation, University of Minnesota, 1963.

WEISS, D. J., & DAWIS, R. V. A multivariate prediction technique for problems involving multifunctional predictor-criterion relationships. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, 1, 229-230.

(Received November 16, 1968)

FEEDBACK AND RESPONSE MODE IN PERFORMING A BAYESIAN DECISION TASK¹

DAVID W. MARTIN² AND CHARLES F. GETTYS³

Ohio State University

In a complex decision-making situation Ss received data generated by one of three hypotheses according to specified conditional probabilities. The Ss inferred which hypothesis had generated the data or estimated the probability of each hypothesis given the data. Feedback was given after each trial either as the hypothesis which generated the data or the probability that each hypothesis generated the data calculated by Bayes' theorem. The two response conditions and two feedback conditions were combined factorially with a group of 16 Ss making 200 responses in each condition. The percentage of trials when Ss chose the most probable hypothesis was significantly higher for the groups responding with a single hypothesis than for the probability response groups, and higher for the Bayesian probability feedback groups than for the groups receiving no feedback. The Bayesian probability feedback group also gave probability responses which were much closer to the optimal probabilities than did the no-feedback group.

Considerable research has suggested that in some situations man is suboptimal in processing probabilistic information and arriving at a decision (see Peterson & Beach, 1967, for a survey of the literature). This lack of optimality has usually shown man to be conservative in probability estimation; the decision-maker's estimates are usually less extreme than those of the optimal model, Bayes' theorem (Phillips & Edwards, 1966). A number of studies have succeeded in determining factors which influence this suboptimal behavior. The present experiment asks if man's inferences are influenced by experience within a decision-making environment in which several modes of response are required and in which feedback is provided concerning the appropriateness of these responses.

In a decision-making situation, a decision-maker infers which state has occurred among a number of possible states of the world. After

collecting relevant information about the world, he may be required to respond either by inferring which state has occurred or by an estimate of the probability that the world is in each state. The former case will be referred to as nominal response mode and the latter by probability response mode. A decision-maker might improve his performance with the aid of feedback. Again, at least two possibilities exist for feedback mode. Either he may be informed about which state the world is in or, if posterior probabilities can be determined from Bayes' theorem, he may be informed of the actual probabilities of each state of the world. Thus, feedback mode can either be nominal or Bayesian probability.

In many situations either nominal or probability responses may be used. A course of action usually is determined by the selection of only one event as that which is most likely to have occurred. However, perhaps this nominal type of response could be made more accurately if the decision-maker were required to respond probabilistically and his largest probability were taken to be his nominal choice. Both nominal and probability information regarding the optimality of a decision are also sometimes available. Particularly in a situation where events occur quite frequently, not only is information concerning what actually happened available, but through the use of records and calculations, a Bayesian

¹ The research reported in this paper was carried out at the Human Performance Center and was sponsored by the Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, under Contract No. AF 33(615)-2248 with the Ohio State University Research Foundation.

² Request for reprints should be sent to David W. Martin, Human Performance Center, Ohio State University, 404-B West 17th Avenue, Columbus, Ohio 43210.

³ Now at the Engineering Psychology Laboratory, University of Michigan.

TABLE 1
FOUR $P(D_{j,k}/H_i)$ CONTINGENCY TABLES

| $D_{j,k}$ | H_1 | H_2 | H_3 |
|-----------|-------|-------|-------|
| $D_{1,1}$ | .22 | .51 | .13 |
| $D_{2,1}$ | .51 | .31 | .15 |
| $D_{3,1}$ | .27 | .18 | .72 |
| $D_{1,2}$ | .38 | .21 | .44 |
| $D_{2,2}$ | .19 | .57 | .27 |
| $D_{3,2}$ | .43 | .22 | .29 |
| $D_{1,3}$ | .73 | .23 | .41 |
| $D_{2,3}$ | .13 | .61 | .20 |
| $D_{3,3}$ | .14 | .16 | .39 |
| $D_{1,4}$ | .37 | .49 | .31 |
| $D_{2,4}$ | .52 | .30 | .25 |
| $D_{3,4}$ | .11 | .21 | .44 |

probability solution might also be determined. Where frequentistic data are not available such as in a high-level command-control system where major war is a low frequency event, estimates from experienced experts might be used as probability feedback in place of an actual Bayesian calculation. It would be helpful to know whether the additional probability feedback could be processed by decision-makers to aid them in their decision or to help them improve their performance.

The present experiment was an attempt to determine which types of response and feedback modes produce responses which most closely approximate Bayesian responses in a given probabilistic environment, and whether this performance improves with practice.

METHOD

Subjects

Sixty-four male university students were paid \$1.25/hr for voluntary participation.

Design

Two response modes, nominal and probability, and two feedback modes, nominal and Bayesian probability, were combined factorially to yield four conditions. Sixteen Ss were assigned randomly to each of the four response-feedback conditions. The Ss under each condition were trained and tested in small groups (6–10 Ss).

Procedure

An abstract decision-making situation was used in which the world could be in one of three equally

likely states called hypotheses, H_i (see Table 1). In each of four contingency tables, called data classes, three possible states could occur with probabilities contingent upon which H_i had been chosen from the three equally likely H_i (e.g., if H_1 were chosen, State 1 would occur in Data Class 1 with a probability of .22, State 2 with .51, and State 3 with .27). To generate the data for a trial the following procedure was used: First, an H_i was randomly selected, and then, in each of the four contingency tables, a data state was chosen for each data class according to the conditional probabilities associated with the H_i selected. These probabilities may be designated $P(D_{j,k}|H_i)$, the probability of a particular data state (j) for each data class (k) given the occurrence of an hypothesis (i).

On each trial Ss were informed which data states had been generated within each data class. Based upon this information, the $P(D_{j,k}|H_i)$ contingency tables, and the knowledge that the H_i were equally likely prior to any information about data states, Ss were required to make one of two types of responses. Under nominal response mode, Ss made a subjective selection of which H_i generated the data states, and under probability response mode, Ss estimated how likely each H_i was, given the data states which had occurred, $\psi(H_i|D)$. After each S had made his subjective response, feedback was given. Under the nominal feedback mode, Ss were told which H_i had generated the data states, and under the Bayesian probability feedback mode, Ss were given the optimal probability estimate computed from Bayes' theorem:

$$P(H_i|D) = \frac{P(H_i) P(D|H_i)}{\sum_i P(H_i) P(D|H_i)}$$

where $P(H_i)$ is the probability of H_i prior to information about data states, $P(D|H_i)$ is computed from the four contingency tables, $P(H_i|D)$ is the posterior probability of H_i given the data, and D is the occurrence of a particular sample of data. As conditional independence was assumed, $P(D|H_i) = P(D_{j1}|H_i) P(D_{j2}|H_i) \dots P(D_{j4}|H_i)$. The Ss were not, however, permitted to do any written calculations, and were not told how to calculate Bayes' theorem. Their estimates, therefore, are assumed to be subjective opinions.

All Ss received a 1-hr. instruction period consisting of a frequentistic definition of probability and a description of the task situation in terms of an objects-and-urns paradigm (H_i 's were urns, dimensions of objects were data classes, levels of each dimension were data states). Five sample trials with feedback were then given to familiarize Ss with the type of decisions to be made and the mechanics of the task.

Data and feedback were presented by means of closed-circuit TV. A display first appeared informing S which data state had occurred in each of the four data classes. For example, S would see a sheet of paper on his monitor which might list Data Class 1 as being in State 2, Data Class 2 in State 1, Data

Class 3 in State 1, and Data Class 4 in State 1. He could then refer to his four contingency tables which were always available and pick out the four appropriate $P(D_{jk}|H_i)$ rows. Each S then made either an estimate as to which H_i generated the sample of data states or else a $\psi(H_i|D)$ for each H_i , depending upon the response condition in which he was operating. These responses were written on the face of an IBM card. The S s in the probability response conditions made three-digit probability responses having values between .998 and .001. In the above example an S making a nominal response might write down H_1 and an S making a probability response might write down .600, .250, and .150, respectively, for each H_i (the actual probabilities in this example being .769, .108, .123). After all S s had deposited their cards in a box in front of them, feedback for that trial appeared on their monitors. The feedback informed them either which H_i had generated the data on that trial or the $P(H_i|D)$ for each H_i on that trial, depending on the feedback condition in which they were operating. A new trial then appeared in the same manner for 200 trials. This required four or five 2-hr. sessions depending upon the self-pacing of the group of S s.

RESULTS AND DISCUSSION

Bayesian Choice Analysis

Since all groups did not make the same type of response, for a comparison between all four groups the probability responses in two of the groups were transformed to nominal responses using the following assumption: S s would have chosen the hypothesis receiving their highest probability had this been required of them. Thus, the dependent measure used to compare the four groups was the percentage of times that S s explicitly or implicitly chose the hypothesis which was most likely from the Bayesian calculations.

When the percentage Bayesian choice data were plotted for the four groups over blocks of 50 trials, significant block differences were found. To determine whether this block effect indicated that learning occurred, a linear component trend analysis of the main effect of blocks was computed. No significant linear component of trend was found, $F(1/240) < 1.00$, and since there was no interaction of blocks with either response mode or feedback mode, no learning would seem to have occurred in any of the groups. The block differences were apparently due to random variation of problem difficulty within blocks rather than learning. For this reason, the

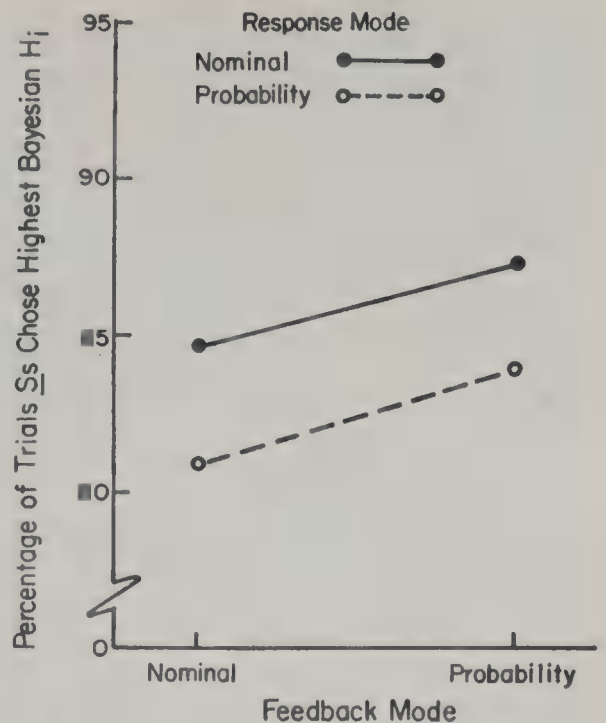


FIG. 1. The percentage of trials on which S s chose the H_i with the highest Bayesian probability as a function of response mode and feedback mode.

percentages were averaged over all trials, and one point was used to describe the data for each group. Figure 1 indicates that both response mode and feedback mode had an effect on performance. The mean difference between the two groups making nominal responses and those making probability responses was 3.6%, and the difference between the two groups receiving nominal feedback and those receiving probability feedback was 2.9%. An analysis of variance of the percentage Bayesian choices indicated that main effects due to response mode, $F(1/240) = 19.09$, $p < .01$, and feedback mode, $F(1/240) = 10.31$, $p < .01$, were significant; no significant interactions were found. The fact that these significant effects were small is also indicated by the proportion of variance accounted for. Estimated ω^2 for response mode was .06 and for feedback mode was .03.

While the differences between groups are significant, the differences might be considered small in terms of the overall level of performance since S s endorsed the same H_i as Bayes at least 80% of the time. The importance of this difference would, of course, depend upon the cost of an incorrect choice.

An evaluation of S 's performance in terms of this discrete-choice measure would thus indicate that superior performance is achieved when S s make nominal responses and probability feedback is given.

Originally it was thought that the use of the probability response mode would cause S s to exhibit a more exacting type of inference and thereby improve their inferred nominal response performance. The results indicate that the opposite was true. The task of stating probabilities reduced the percentage of occasions when S 's highest probability estimate was under the hypothesis receiving the highest Bayesian estimate. A possible explanation for this unexpected result would be that probability responses require different information-processing behavior than nominal responses require. When nominal responses are made, for example, perhaps only the few hypotheses judged to be most likely need be considered since S 's task is simply to choose the most likely hypothesis. If S 's response is a probability, he should be concerned with the likelihood of all the hypotheses.

The finding concerning response mode would thus indicate that if the output required of a decision-maker is simply a choice of the most likely hypothesis, his performance is better with a nominal response than it is with a probability response which involves the processing of more information. This would seem to be true regardless of the feedback given.

The S s receiving probability feedback were superior to those receiving nominal feedback in terms of the choice measure. The probability feedback may have given S s a better idea of how probabilities ought to be aggregated or interpreted. A second possibility is that the probability feedback groups actually received more valid information regarding optimal performance than the nominal feedback groups. The S s receiving probability feedback always knew which hypothesis was most likely, but those receiving nominal feedback were only informed which hypothesis generated the data. The generating hypothesis does not always have the largest Bayesian probability, and for this reason is not always the optimal nominal response. The hypothesis that generated the data will be different from

the hypothesis having the largest posterior probability with a probability of $1 - P(H^*|D)$, where H^* is the hypothesis which generated D . Thus, the nominal feedback groups did not always receive information regarding optimal response, while the probability feedback groups did. To determine which of these explanations is true in a future experiment, a nominal feedback condition could be added in which the hypothesis with the highest Bayesian probability is given as feedback rather than the generating hypothesis.

D Score Analysis

By using a second dependent measure, the estimates of the two probability response groups could be compared to the probabilities calculated from Bayes' theorem (Schum, Goldstein, Howell, & Southard, 1967). This measure, D , is based upon the difference between the log of the likelihood ratio inferred from S 's response and the likelihood ratio from the calculation of Bayes' theorem;

$$D = \log \frac{\psi(H^*|D)}{1 - \psi(H^*|D)} - \log \frac{P(H^*|D)}{1 - P(H^*|D)}$$

The difference measure (D) thus is an expression of the correspondence between S 's response and the optimal response. When these responses are equal S 's inferred likelihood ratio and the Bayesian likelihood ratio are the same and the difference between them is 0; $D = 0$. A positive D indicates that S responded with a probability under H^* larger than that calculated from Bayes' theorem, and a negative D indicates that S 's probability was smaller (conservatism).

To compare the probability response groups, an analysis of variance was performed on the D scores. A nonsignificant block effect, $F(3/120) = 1.08$, $p > .25$, indicated that performance was essentially flat across trials with no indication of learning. There were, however, significant main effects due to feedback condition, $F(1/120) = 40.19$, $p < .01$. When using D score, which is a much more sensitive measure than percentage correct choices, the proportion of variance accounted for by feedback mode was .24. The mean of the D scores for nominal feedback was $-.25$ and for probability feedback, $-.11$. Since perfectly Bayesian behavior would yield a D of 0, both

groups of Ss could be considered to have responded with probabilities less extreme than those indicated by Bayes' theorem, but the group receiving probability feedback responded significantly closer to optimality than did the nominal response group.

Since the diagnosticity of an item or a sample of items has been found to be a major determinant of performance in past experiments, the responses in the present experiment were grouped into three diagnosticity categories. Diagnosticity was not a major variable in the experiment, so it was not included in a factorial manner. The grouping into diagnosticity levels was accomplished by ordering the 200 trials by the size of the Bayesian $P(H|D)$ under the hypothesis which generated the sample, and then separating these trials into three categories. The $P(H|D)$ under the generating hypothesis in each diagnosticity category were as follows: low, .058-.434; medium, .439-.770; and high, .781-.954. The number of trials per block representing each diagnosticity level was not equal. The mean D scores for the two groups responding with probabilities are shown in Figure 2. An inspection of Figure 2 shows no indication of learning; yet there are obvious differences in behavior for the two groups. The data from both groups would seem to support past research indicating less conservative and even excessive estimates when item diagnosticity is low, and more conservative estimates when diagnosticity is high (Peterson & Miller, 1965). However, the performance for the group receiving probability feedback was, at all three diagnosticity levels, much closer to optimal than the corresponding performance for the nominal feedback group. In fact, the differences in divergence from optimality between the two groups was larger than the single means across diagnosticity levels previously reported would indicate, since suboptimal behavior above zero tended to make suboptimal behavior below zero appear more optimal.

No evidence of learning was found when the data were analyzed in 50-trial blocks, even though there were other significant main effects. An analysis of the first 50-trial block was undertaken to determine if learning took place in the first 10 to 50 trials (rapid learn-

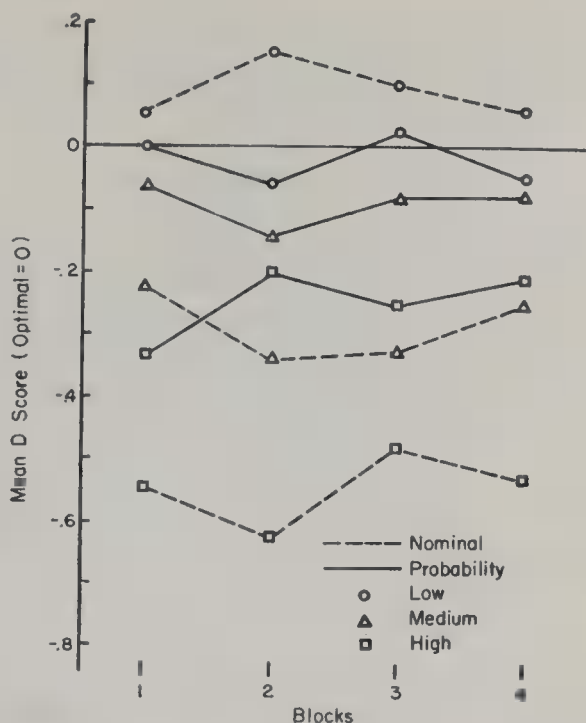


FIG. 2. Mean D score for probability-response groups as a function of feedback mode (nominal and probability) and sample diagnosticity (high, medium, and low).

ing). Again, no clear-cut trend effects were found. Both dependent measures, however, were so sensitive to between-trial diagnosticity differences that any rapid learning effects might have been buried in this variability. Rapid learning might also have occurred during the five preexperimental practice trials.

As an attempt to determine whether rapid learning effects could be found, the probability-response, probability-feedback condition was replicated using 16 different Ss. However, only 20 trials were used and problems were randomized within these trials. This group performed at about the same D score level as the previous group, but again no short-term trend effects were found.

Conclusions

Thus, in the probability response groups, whereas the type of feedback had small but significant effects in terms of hypothesis choice, it had large effects in terms of the appropriateness of the size of the probability estimates. Figure 2 indicated that the groups receiving only nominal feedback behaved in a much less optimal manner than the prob-

ability feedback group. The D scores for the nominal feedback group were at least twice the size of those for the group receiving probabilities for feedback.

These results concerning feedback would indicate that even in situations in which a decision-maker is simply required to choose a most likely hypothesis, his performance can be enhanced by presenting probability feedback. A trade-off undoubtedly exists here, however. The improvement in choice performance is small, even though significant, and the cost of determining a Bayesian probability vector (assuming that it can be determined) is likely to be high in most situations. However, in a situation where a probability response is required, probability feedback seems to be very important if S is to make estimates that are at all close to Bayesian estimates. This is true particularly when evidence is highly diagnostic. In some situations, determining Bayesian probabilities is an impossibility, but in those situations where it is possible, particularly when probability responses are required, Bayesian probability feedback should be given.

Perhaps the only appropriate means of implementing probability feedback in many situations where posterior probabilities are unavailable would be by training decision-makers in a situation where veridical posterior probabilities are available and transferring them to the situation where they are not. The present study does not, of course, answer questions concerning transfer, but it would seem that improved performance under probability-response and probability-feedback conditions might transfer to conditions without feedback.

REFERENCES

- PETERSON, C. R., & BEACH, L. R. Man as an intuitive statistician. *Psychological Bulletin*, 1967, **68**, 29-46.
- PETERSON, C. R., & MILLER, A. J. Sensitivity of subjective probability revision. *Journal of Experimental Psychology*, 1965, **70**, 117-121.
- PHILLIPS, L. D., & EDWARDS, W. Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 1966, **72**, 346-354.
- SCHUM, D. A., GOLDSTEIN, I. L., HOWELL, W. C., & SOUTHARD, J. F. Subjective probability revisions under several cost-payoff arrangements. *Organizational Behavior and Human Performance*, 1967, **2**, 84-104.

(Received November 14, 1968)

INTERFERENCE BETWEEN CONCURRENT TASKS OF DRIVING AND TELEPHONING¹

I. D. BROWN,² A. H. TICKNER, AND D. C. V. SIMMONDS

Applied Psychology Research Unit, Cambridge, England

Twenty-four men were given the task of judging whether to drive through gaps which might be larger or smaller than the car. They were also given a telephoning task of checking the accuracy of short sentences. Interference between the concurrently performed tasks was investigated. Telephoning mainly impaired judgments of 'impossible' gaps ($p < .01$). The control skills employed in steering through 'possible' gaps were not reliably degraded, although speed of driving was reduced ($p < .01$). Driving increased errors ($p < .01$) and prolonged response times ($p < .005$) on the sentence-checking task. It is concluded that telephoning has a minimal effect on the more automatized driving skills, but that perception and decision-making may be critically impaired by switching between visual and auditory inputs.

In the next decade a substantial increase is expected in the number of radiophones fitted to road vehicles. The user population will probably include many more car drivers on business trips who may be handling far more complex messages than those transmitted by present professional drivers employed in the police, fire, ambulance, and taxi services. The question arises as to whether this concurrent activity will impair driving skills sufficiently to increase the risk of accident on the road. To the authors' knowledge there is no direct evidence from research or from accident statistics which answers this question conclusively, although there is limited evidence that attention to auditory stimuli has little effect on the control skills employed in driving (Brown, 1965, 1966, 1967). This paucity of reliable information has led to the anomalous situation, in the United Kingdom at least, in which a driver is permitted to telephone but

risks prosecution if he performs some more automatized task, such as shaving with an electric razor.

It is clear that there are two main sources of interference between driving and telephoning. Having to use a hand microphone and having to manipulate push buttons to make or take a call will be inconvenient and may impair steering, gear changing, or other control skills. This is a problem which may be solved by engineering advances and is not the concern of the present paper. A more important and lasting problem arises from the hypothesis that man can be considered to act as a single communication channel of limited capacity. The prediction from this hypothesis is that the driver will often be able to telephone only by switching attention between the informational demands of the two tasks. Telephoning could thus interfere with driving by disrupting visual scanning, since visual and auditory information would have to be transmitted successively. It could also interfere by overloading short-term memory and impairing judgment of relative velocity, which depends upon integration of successive samples of visual information. Less important, intermittent sampling of the telephone message could result in partial or complete failure of communication, depending upon the redundancy in the message.

The object of the present experiment was to investigate this effect of divided attention

¹The authors would like to thank the United Kingdom Ministry of Transport for providing some of the apparatus, most of Ss, and two research assistants. The General Post Office kindly provided the radiophone equipment and a technical assistant. The experimental car and financial assistance were supplied by the Medical Research Council. Thanks are also due to the Ministry of Defence (Air) for permitting part of an airfield to be used as a test track. The research was carried out under the general direction of R. Conrad.

²Requests for reprints should be sent to Ivan D. Brown, Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

on judgments of clearance, control skills, and the checking of auditory messages.

METHOD

Driving Task

Four different aspects of driving skill were measured:

(1) The ability to judge whether a gap between two obstacles was wide enough to be cleared by the car. The gap was set up at right angles to the track and could be varied from 3 in. smaller than the car to 9 in. wider, in steps of 3 in. There were thus three 'possible' gaps and two 'impossible.' This range of sizes was chosen on the results of a pretest which showed that Ss thought it possible to drive through a gap 3 in. wider than the car on about 50% of the trials.

Each trial drive required S to make judgments on 20 gaps (4 of each size) which were arranged in random order and spaced at equal intervals around a circuit 1.5 ml. in length. The S was not allowed to stop during a trial, but otherwise driving was self-paced. If he drove past the marker placed 18 ft. in front of each gap he was considered to have accepted the gap as possible and no other indication of his decision was required. Having accepted a gap he had to drive straight through it, regardless of its actual size. He indicated rejection of a gap by turning left at the marker and rejoining the track on the far side of the gap. This alternative route was designed to impose a comparable delay for accepted and rejected gaps, in order to minimize any biasing of the data by Ss who might be motivated to maintain speed instead of attempting to drive through the more difficult gaps.

(2.) A second measure of driving performance was obtained by recording the number of possible gaps which were cleared successfully by S when he decided to drive straight ahead. He was considered to have failed if any part of the car touched either obstacle.

(3.) Third, speed of performance was obtained by recording the time taken to drive around each complete circuit of 20 gaps.

(4.) Finally, a set of measures was recorded automatically from the frequency with which S used the steering wheel and foot controls of the car and from the lateral and longitudinal accelerations he imposed on it.

Telephoning Task

Messages transmitted over the radiophone presented S with a reasoning test based on grammatical transformation, described in detail by Baddeley (1968). The S heard a series of sentences, each of which was followed by the letters "A" and "B." Each sentence claimed to describe the order in which the following pair of letters would be spoken and S had to decide whether the description was true or false.

He responded by speaking the words "True" or "False," unless he had missed the message, in which case he asked for the next sentence.

Examples of the sentences are as follows:

| Examples: | Ss Correct response |
|---------------------------|---------------------|
| A follows B—BA | True |
| B precedes A—AB | False |
| A is followed by B—BA | True |
| B is not followed by A—BA | False |
| B is preceded by A—BA | False |
| A does not precede B—BA | True |

As in Baddeley's initial experiment, the test comprised all 64 possible combinations of the following 6 binary conditions: (a) positive or negative, (b) active or passive, (c) true or false, (d) precedes or follows, (e) A or B mentioned first in the sentence, (f) following letter pair AB or BA used. Five different test forms were used, with the 64 items occurring in a different random order on each. It took about 2.5 sec. to read out each sentence and letter pair, and S had to respond as quickly as possible, consistent with accuracy. There was a pause of about 2 sec. between S's response and the beginning of the next sentence.

Telephoning performance was measured by scoring each response as correct or incorrect and by recording the time taken to respond after the letter pair had been spoken.

Apparatus

All tests were run in an Austin A40 estate car, which had manual gearshift, braking, and steering. The car was 5 ft. wide. Frequency of control movements was recorded on digital counters enclosed in a soundproof box. They were operated via micro-switches which were activated whenever the steering wheel, accelerator, brake, or clutch pedals were moved sufficiently to change the velocity of the car. Three other counters recorded lateral accelerations by pooling positive and negative readings within the ranges .1-.2 g, .2-.3 g, and greater than .3 g. Longitudinal accelerations were similarly recorded. Counter readings were obtained photographically before and after each trial. Driving time per circuit was measured by stopwatch.

The telephoning task was presented by E₁ from a mobile transmitter truck parked beside the track. Messages were received in the car from a loudspeaker mounted in front of S and he responded via a telephonist's headset. The E₂ sat at a small console in the back of the car and controlled the transmit/receive selector of the telephone link by footswitch. Thus S had no radiophone controls to manipulate and any impairment of driving skills could be attributed to divided attention. The E₁'s transmissions and S's responses were recorded on magnetic tape for subsequent analysis of response times.

The gaps were formed by pairs of obstacles 4 ft. high and 20 in. wide, constructed from hardboard on softwood frames and painted white.

TABLE 1
EFFECT OF TELEPHONING ON ERRORS
OF GAP-JUDGMENT

| Clearance in gap (in.) | % errors when driv- ing alone | % errors when tele- phoning concur- rently | Increase in % errors | <i>p</i> ^a |
|------------------------------|-------------------------------------|--|----------------------------|-----------------------|
| -3 | 28.0 | 47.2 | 19.2 | < .01 |
| 0 | 70.8 | 93.0 | 22.2 | < .01 |
| 3 | 79.5 | 81.2 | 1.7 | > .05 |
| 6 | 28.5 | 39.2 | 10.7 | > .05 |
| 9 | 7.7 | 18.5 | 10.8 | = .05 |

Note.—Data have been corrected for guessing.

^a One-tailed Wilcoxon matched-pairs signed-ranks test of significance, (see Siegel, 1956, p. 75).

The Ss were 24 men within the age range 21–57 (median age 41). Their car-driving experience, as judged from the length of time they had held a license, ranged from 3 to 37 yr. (median time 15½ yr.). Twenty-two men were volunteers from various establishments of the United Kingdom Ministry of Transport, the remaining 2 were drawn from the APRU research panel. Only 1 S had experience on mobile radiophones.

Procedure

After a short introductory explanation of the experimental objectives and method, S had 5 min. practice on sentence-checking alone in the stationary car, followed by one practice trial of driving alone during which he made judgments of the gaps and tried to drive through those which he considered were possible. Finally, he had one practice trial of driving and telephoning concurrently.

This was followed immediately by six test trials: three in the order given during practice and three in the reverse order. Thus each S had 10 min. of testing on sentence-checking alone and two trials of driving alone. These provided individual baseline measures of performance, for comparison with the two trials of driving while telephoning.

There was always a short pause between successive trials of driving, during which the size of the gaps was altered according to a prearranged schedule. Thus S met the various sizes of gap in a different order on each trial.

RESULTS

Learning on the tasks of driving and telephoning was negligible. Therefore data from the first and second trials in each condition were pooled to give overall measures of performance on telephoning alone, driving alone, and telephoning while driving.

Table 1 shows that errors of gap-judgment were higher for all sizes of clearance when Ss also had to telephone, although the difference was statistically significant only for the two impossible gaps and for the largest possible gap. This means that when Ss operated under conditions of divided attention, they tried to drive through far more gaps that were smaller than the car and slightly fewer that were larger.

Table 2 shows that skill in steering through possible gaps was not reliably impaired by telephoning, although there was a tendency for performance to be degraded when clearance was reduced to 3 in.

Driving time per circuit was 361.3 sec. when Ss were driving alone and 385.2 sec. when they were telephoning concurrently. This 6.6% reduction in speed was statistically significant ($p < .01$, Wilcoxon test). Telephoning had no reliable effect on the frequency with which the car controls were used, nor did it affect the lateral and longitudinal accelerations imposed on the vehicle ($p > .05$, Wilcoxon test).

Table 3 shows that speed and accuracy of telephoning performance were both substantially impaired when Ss also had to drive.

It can be inferred from the observed reduction in speed when Ss had to telephone concurrently that they were attempting to gain time in which to handle the additional informational load. However, Tables 1 and 3 show that this change in speed was insufficient to prevent mutual interference between gap-judgment and sentence-checking. It seemed

TABLE 2
EFFECT OF TELEPHONING ON SKILL IN STEERING
THROUGH POSSIBLE GAPS

| Clearance in gaps (in.) | % gaps cleared when driv- ing alone | % gaps cleared when tele- phoning concur- rently | Change in % gaps cleared | <i>p</i> ^a |
|-------------------------------|--|---|--------------------------------|-----------------------|
| 3 | 84.1 | 75.8 | 8.3 fewer | > .05 |
| 6 | 95.2 | 95.5 | .3 more | > .05 |
| 9 | 99.2 | 99.3 | .1 more | > .05 |

^a Wilcoxon test.

TABLE 3
EFFECT OF DRIVING ON TELEPHONING
PERFORMANCE

| Measure of performance | Sentence-checking alone | Sentence-checking when driving concurrently | Change in performance measure | <i>p</i> ^a |
|-----------------------------------|-------------------------|---|-------------------------------|-----------------------|
| % errors (corrected for guessing) | 23.8 | 45.0 | 21.2 more | <.01 |
| Response time (sec.) | 1.81 | 2.60 | .79 longer | <.005 |

^a Wilcoxon test.

interesting to look further into the way in which the additional driving time was used under conditions of divided attention. This was done by calculating for each *S*: (*a*) the change in driving time and in gap-judgment errors from driving alone to driving while telephoning, and (*b*) the change in errors and in response-time on sentence-checking from telephoning alone to telephoning while driving.

There was a significant positive correlation between increase in driving time and increase in errors of gap-judgment ($\tau = .314$, $p < .0316$; Kendall's rank correlation coefficient, see Siegel, 1956, p. 213). There are three possible explanations of this correlation: (*a*) that judgment of clearance was impaired by reduced speed per se, which seems improbable, (*b*) that the increased time resulted from the greater caution with which *Ss* drove through the additional impossible gaps they had judged incorrectly when telephoning, (*c*) that the increased time was taken in order to maintain performance on the telephoning task, which biased attention away from the task of gap-judgment. The latter explanation is supported by the finding that increase in driving time and increase in errors on sentence-checking were negatively correlated ($\tau = -.441$, $p < .0026$). The correlation between increase in driving time and increase in response time on sentence-checking was small and statistically unreliable ($\tau = .074$, $p < .230$). Therefore the inference is that *Ss* were using the additional driving time to maintain accuracy of sentence-checking. However, Explanation 2 above is also tenable, suggesting

that the positive correlation between increase in driving time and increase in errors of gap-judgment may, at least partly, have been an artifact of the experimental procedure which required *Ss* to drive through all accepted gaps, regardless of their actual size.

There was a suggestion that the order of priority given to the two concurrent tasks was a function of age. Younger *Ss* tended to maintain performance on sentence-checking at the expense of increased errors on gap-judgment. The reverse tendency was observed among older *Ss*, although the effect of telephoning on driving time was comparable for all age groups. With the present small group of *Ss* these age effects were statistically unreliable ($p > .05$; Kruskal-Wallis one-way analysis of variance, see Siegel, 1956, p. 184).

DISCUSSION

Judgments of clearance were degraded across the complete range of gaps used in the experiment, although the effect was small on gaps larger than the car (see Table 1). There are three possible sources of this degradation: (*a*) interference between representations of visual and auditory stimuli within short-term memory, which would have impaired integration of successive visual samples of gap size. As successive sampling is not essential to judgments of clearance, this seems an unlikely source of the observed interference. (*b*) a relaxation of the criteria on which gap judgments were made. Although this explains the increase in errors on impossible gaps, it could not have been the sole source of interference, or more gaps of any size would have been accepted when *Ss* were driving and telephoning concurrently. In fact, *fewer* possible gaps were accepted under this condition (see Table 1). (*c*) an impairment of perception resulting from switching between sensory modes. This also could not have been the sole source of interference, or the most difficult judgment (of 3 in. clearance, see Table 1) would probably have been degraded most rather than least. The possibility that a ceiling effect was operating to stabilize errors at this level is ruled out by the finding that telephoning had a greater impact on judgments of zero clearance.

It must be concluded that concurrent telephoning produced both a relaxation of criteria and an impairment of perception. During judgments of possible gaps, the tendency for impaired perception to produce errors of rejection would have acted in opposition to the tendency for relaxed criteria to produce errors of acceptance. During judgments of impossible gaps, both impaired perception and relaxed criteria would have produced errors of acceptance. This explanation would account for the finding that divided attention had the differential effect of causing a significantly large increase in acceptance of impossible gaps, but a smaller and mainly insignificant increase in rejections of possible gaps. Since the latter effect was an *increase* in rejections, it may be inferred that performance on gap-judgment was affected more by impaired perception than by relaxed criteria, assuming that the relative importance of the two effects was stable across the range of gaps used.

The perceptual-motor skills employed in steering through possible gaps were not reliably affected by telephoning (see Table 2). Performance was beginning to deteriorate when clearance was reduced to 3 in., but this source of interference is unlikely to present a major problem on the road, since even in the experiment Ss were rejecting about 80% of these gaps as impossible (see Table 1). Steering skills are probably so automatized among trained drivers that they are minimally degraded when attention has to be diverted intermittently to an auditory stimulus. However, it must be noted that drivers currently have to manipulate equipment to make and take calls, and these motor activities could interfere with steering skills as well as with the use of other manual controls.

The increase in driving time observed when Ss were telephoning could have resulted simply from the greater caution with which they drove through the additional impossible gaps they had judged incorrectly in this condition. This would account for the finding that errors of gap-judgment were positively correlated with driving time. The alternative explanation is that Ss deliberately reduced speed in order to handle the additional load. If the latter were true, the results indicate that the

additional time was used to maintain performance on the telephoning task, at the expense of increased errors on gap-judgment. It is impossible to distinguish conclusively between these explanations on the experimental data alone, because driving and telephoning were temporally independent and because driving time was measured only as a total score per trial. Further research would be needed to investigate these alternatives and to study the possibility that order of priority given to concurrent tasks of driving and telephoning may be a function of the driver's age. It would also be necessary to investigate the *timing* of control skills before it could be concluded that these are entirely unaffected by divided attention.

Both speed and accuracy of telephoning were affected by the driving task (see Table 3). It is impossible to say how much of this decrement simply resulted from concurrent performance of the usual control skills employed in driving and how much resulted from the experimental tasks of judging clearance and driving through gaps, because conditions of testing precluded any detailed recording of the temporal relationships between driving and telephoning. It seems clear that calls via mobile radiophones will take longer than ordinary calls, since Table 3 shows that messages containing little redundancy are substantially affected by driving, therefore repetitions would be necessary in order to transmit all the information in practice. Even with the greater redundancy of plain speech messages, complete failure of communication could occur if division of attention were dictated by traffic conditions rather than by the content of the telephone message.

The general conclusion must be that some mutual interference between the concurrent tasks is inevitable under conditions of telephoning while driving on the road. The results suggest that, although more automatized control skills may be affected minimally by this division of attention, some perceptual and decision skills may be critically impaired. The extent to which this impairment is a function of the driving task, the informa-

tional content of the telephone message, and the individual characteristics of the driver must remain a subject for further research.

REFERENCES

- BADDELEY, A. D. A 3 minute reasoning test based on grammatical transformation. *Psychonomic Science*, 1968, **10**, 341-342.
- BROWN, I. D. Effect of a car radio on driving in traffic. *Ergonomics*, 1965, **8**, 475-479.
- BROWN, I. D. Effects of prolonged driving upon driving skill and performance on a subsidiary task. *Industrial Medicine and Surgery*, 1966, **35**, 760-765.
- BROWN, I. D. Measurement of control skills, vigilance, and performance on a subsidiary task during 12 hours of car driving. *Ergonomics*, 1967, **10**, 665-673.
- SIEGEL, S. *Nonparametric Statistics*. New York: McGraw-Hill, 1956.

(Received October 21, 1968)

COMPARISON OF "REAL LIFE" ASSESSMENTS OF MANAGEMENT POTENTIAL WITH SITUATIONAL EXERCISES, PAPER-AND-PENCIL ABILITY TESTS, AND PERSONALITY INVENTORIES

J. R. HINRICHS¹

IBM, White Plains, New York

The major component of evaluation from a 2-day assessment program covering 47 members of a large national marketing organization consisted of ratings of degree of active participation in the group situational exercises, followed by ratings of administrative and decision-making ability. Paper-and-pencil ability tests and personality inventories were less clearly related to assessments of managerial potential. Ratings of management potential developed from a careful review of company personnel records were as highly correlated with the assessment center data as were overall ratings from the 2-day program, except for ratings dealing with interpersonal behavior.

One of the most pressing problems which industry will face over the coming decade is a severe shortage of qualified management personnel. The combined effects of significant expansion in business activity along with a labor force in which there will be a decline in the absolute numbers of participants in the key manager age bracket of 35-45 are more and more forcing companies to take a hard look at the procedures which they utilize for the identification of management talent. As a result, the current focus is on the *early* identification of talent, as highlighted by the classic study undertaken by the Standard Oil Company of New Jersey (Laurent, 1961).

Increasingly, a variety of assessment techniques are being utilized in an effort to predict managerial potential, and studies are showing that a spectrum of inputs can add valid information to the old standbys of group tests of ability and temperament. Although previous research has often reported negative findings, recent research suggests that the interview can be a valid assessment procedure if carefully conducted (Ghiselli, 1966; Prien, 1962). And, contrary to the assumptions of many industrial psychologists, clinical techniques have been shown to have promise (Albrecht, 1964).

An approach which has been used increasingly involves situational tests (Flanagan, 1954) in an effort to approximate closely the

kinds of behaviors required for managerial jobs. A number of companies have been experimenting with systematic management assessment programs built largely around situational exercises (Bray & Campbell, 1968; Bray & Grant, 1966; Grant, Katkovsky, & Bray, 1967; Greenwood & McNamara, 1967; Hardesty & Jones, 1968). Such programs are designed to bring potential candidates for management together for several days, to have them participate in a number of group and individual exercises, to have them take a battery of ability and personality tests, and then to have a team of observers distill the results of the program into a series of predictions of probable management potential for each of the program participants.

This is clearly a costly process. In addition to the facilities and materials required to run such a program, even greater costs of several days attendance by the participants and by the staff of observers suggests that management invests a great deal of money into the assessment center process in order to arrive at a relatively straightforward prediction of promotability. While research has shown, at least at AT&T, that assessment evaluations are predictive of subsequent promotion into management and of salary growth (Bray & Grant, 1966), we are still left with several rather gnawing questions:

(1) How much does the program contribute beyond what is already known about the candidates? In one sense the answer is "Nothing," since typically the effectiveness of the

¹ Requests for reprints should be sent to the author, John R. Hinrichs, IBM Corporation, 112 East Post Road, White Plains, New York 10601.

assessment process is subsequently validated against what happens in the organization under more naturalistic conditions, that is, whether or not in the normal course of events an individual is promoted. If this is the ultimate criterion against which an assessment program is validated, one might question "Why bother with the program in the first place?" The answer, of course, is that one hopes the assessment program will be able to identify promotable people earlier in their careers, that it will help to clarify some of the skills important in promotion, and that it will perhaps identify some people who should be promoted but who might under normal circumstances be overlooked. Hopefully, also, the program will fulfill to some extent a personal development function by providing practice in group situations, individualized feedback regarding observed strengths and weaknesses, and greater understanding regarding the caliber of the competition participants are up against. But where is the trade-off, no matter how valid the assessment process, between the use of an assessment program and the prevalent approach to managerial selection used in most organizations of "letting the chips fall where they may"?

(2) A parallel question concerns how much is gained by the use of situational tests—the group and individual exercises which really are the most costly aspects of the assessment center approach—beyond what can be obtained by more traditional paper-and-pencil tests. And if an additional input is obtained from situational exercises, is it an increment sufficiently large to justify the cost?

(3) Another question asks how much redundancy there is in the data collected in a typical assessment program: How many of the tests and exercises overlap, and to what extent can the program be streamlined in an effort to make it both more efficient and less expensive?

The purpose of the research reported in this paper was to evaluate preliminary data from one management assessment program from a number of points of view. In any evaluation of this type, of course, the most pressing problem is the lack of criteria. The nature of the program calls for a predictive validation strategy, since this is what the

program is all about: the prediction of management potential. And this is what the AT&T research has focused upon, though even here the full assessment of validity was felt to be a bit premature and the predictive data were allowed to mellow for only 8 as opposed to the originally intended 10 yr. (Bray & Grant, 1966). In the present instance, it was necessary to obtain some evaluation of the program's value before a 10- or 15-yr. delay, and toward this end this research was designed to relate the various components of the program to several criteria of value. Although this is far from an adequate validation, and purely concurrent in nature, it was hoped that the analysis would provide some basis for evaluation by looking at the relationship between the program prediction and (a) *external criteria*, in the form of some assessment of current value to the organization of the program participants, (b) *internal criteria*, in the sense of relationship between the program components and the overall management potential evaluation flowing from the 2-day program, and (c) *parallel criteria*, in the sense of presently available evaluations of managerial potential which were representative of those utilized in the normal promotional system of the organization.

The research was also designed to determine what makes up the final evaluation—what gets the major weight in the assessment prediction in this particular organization, and to what extent there is redundancy in the measures which are collected. At the same time, an evaluation of the extent to which the situational and the paper-and-pencil methodologies yielded comparable assessments could be viewed as a rough approach to obtaining some degree of construct validation of the concepts covered in the program. Another goal of this analysis was to determine how to reduce meaningfully the number of variables collected so as to understand better the relationships among them and to facilitate analysis of the data.

DATA COLLECTION

The Ss in this study were 47 college-educated male employees engaged in marketing activities for a large technology-based organi-

zation. The Ss came from all areas of the country and had been employed by the organization for an average of 7.3 yr. at the time of assessment. Their average age was 33. Some Ss were recently promoted firstline managers, and some were nonmanagerial personnel who had been identified by management as having management potential. They were selected at random within the criteria of being either new managers or senior personnel considered promotable in the very near future to a management position. Groups of managers and nonmanagers were kept separate for the various situational exercises in the program.

The assessment program employed was very similar to that described by Bray and Grant (1966). Group situational exercises were identical to those described by Greenwood and McNamara (1967): (a) Leaderless Group Discussion, (b) Task Force Committee, and (c) Manufacturing Game. Individual situational exercises included (a) an In-Basket of 25 items which the individual was given an hour and a half to cover. He was then interviewed by a member of the assessment staff regarding his decisions and rated on this performance. (b) A Stock Market Exercise, in which the individual had to respond to hypothetical market fluctuations investing a certain sum of money. (c) A Job Environment Report in which the individual was requested to describe his job in narrative form.

Paper-and-pencil tests were also given throughout the 2-day program and consisted of (a) Concept Mastery Test, Form T, (b) School and College Ability Tests (SCAT), Form U—(Numerical Part only), (c) Gordon Personal Profile, (d) Allport, Vernon, Lindzey Study of Values, (e) Leadership Opinion Questionnaire, (f) Ghiselli Self-Description Inventory, (g) Risk-Taking Scale (Williams, 1965), and (h) a background and contemporary data questionnaire of personal history questions. Keys for general management potential and for self-confidence had been previously developed, and two scores were derived for each program participant.

At the conclusion of the two days of situational and paper-and-pencil tests, the observers who formed the assessment staff and

who were all operating management personnel at least two levels above program participants met to discuss the various exercises and to consolidate the ratings they had made of the participants on traits which were appropriate for the exercise which they observed. Twelve trait ratings were derived as a result of all of the situational exercises: (a) Aggressiveness, (b) Persuasive or Selling Ability, (c) Oral Communications, (d) Planning and Organization, (e) Self-Confidence, (f) Resistance to Stress, (g) Written Communications, (h) Energy Level, (i) Decision Making, (j) Interpersonal Contact, (k) Administrative Ability, and (l) Risk Taking.

These ratings were on a 5-point scale from 1—outstanding—to 5—definitely below average. In addition, an overall evaluation of management potential, which we have termed our “internal criterion,” was arrived at in this evaluation session based on ratings in the situational exercises and scores on the various tests. Analyses have shown adequate reliability for these types of ratings (Greenwood & McNamara, 1967).

In an effort to obtain a comparable evaluation of management potential based only on data readily available to the organization, two experienced managers were asked to review all of the personnel records for the 47 participants in the program and whatever additional information would normally be available to them in making an initial promotional decision. They then rated the 47 people on management potential utilizing a similar rating scale to that employed in the assessment program. We have termed these ratings a “parallel criterion.”

These two managers were assistants to district marketing managers who quite frequently became involved in the evaluation of qualifications of candidates for promotion in the organization. In effect, the task given them was to perform the type of analysis they would ordinarily go through in their day-to-day job activities in coming up with candidates for promotion to recommend to their district manager. The instructions to them stated

This will be essentially a subjective process of distilling all of the information which is available

on 47 men to arrive at a judgment regarding their long-term potential to the company. By long-term we don't necessarily mean until age 65, but probably out for a period of 10-15 years. After going through all of the materials, record your best judgement of the highest level in the company which you can see this individual attaining, based upon the descriptions given on the attached pyramid of position levels. Proceed as if you were trying to select a candidate for an important promotion.

- a. Work through all the materials in the personnel jacket. Specifically look at educational and experience qualifications, appraisal evaluations, any special commendations or accomplishments.
- b. Review his performance evaluations. In addition to his appraisal ratings, see how he is ranked by management among his peers. Is he on the promotion list? The outstanding employee list? What is his sales record? Rate of earning growth? How was he evaluated in training?
- c. Talk with his immediate manager, and, if desirable, his prior managers. Make liberal use of the telephone. Finally, make a judgement, independently, and record it.

Each of these managers spent over a day independently reviewing the 47 personnel jackets and arriving at their ratings. Although they saw this as a tedious task, they were able to make these predictions. It should be recognized that this parallel criterion was not based upon any face-to-face confrontation with the assessee; the evaluators indicated that they were not personally familiar with the people they were rating,

with the exception of one or two cases, and only rarely did they see fit to utilize the telephone for any kind of references from present or past managers. So their ratings came almost exclusively from their personal assessment of company personnel records. And, needless to say, they had no knowledge of the ratings assigned in the 2-day assessment program.

ANALYSIS AND RESULTS

In an effort to clarify the concepts being evaluated in the exercises and to reduce the number of variables being dealt with to a more manageable number of relatively unique scales, the trait ratings which synthesized the observers' ratings in the situational tests were factor analyzed. For this analysis, the 12 ratings were intercorrelated, the maximum row element was inserted in the main diagonal as an estimate of communality, and the matrix was factored by the method of principal components. Between two and six factors were rotated obliquely using Carroll's (1960) biquartimin rotation. The three-factor solution shown in Table 1 seemed best to represent the structure of this particular matrix.

Factor 1 is clearly an activity factor, with heavy loadings from ratings of persuasiveness, aggressiveness, energy level, interpersonal contact, oral communications, and self-confidence. The relatively pure factor picture is one of active participation in the group situational exercises.

Factor 2 reflects more individually based administrative kinds of skills, with heavy loadings of decision making, planning and organizing, written communications, and administrative ability.

The third factor, while not showing as large loadings as the previous two, appears to reflect a component of resistance to stress with loadings for resistance to stress and risk taking. In large measure this is a residual factor since both of these traits have their primary loadings on Factors 1 or 2. An examination of the rotations of more than three factors for these data does not suggest additional meaningful factors; it would seem that at most three components of evaluation are being derived from the various situational tests with this particular population and this particular

TABLE 1
FACTOR ANALYSIS OF ASSESSMENT RATING SCALES

| Rating scale | Rotated factor loadings | | | |
|-------------------------------|-------------------------|-----|-----|----------------|
| | 1 | 2 | 3 | h ² |
| Persuasive or Selling Ability | 95 | -15 | 04 | 92 |
| Aggressiveness | 89 | 02 | -15 | 89 |
| Energy Level | 82 | -03 | 14 | 72 |
| Interpersonal Contact | 81 | 11 | 02 | 77 |
| Oral Communications | 79 | 13 | 08 | 75 |
| Self-Confidence | 73 | 23 | 12 | 75 |
| Decision Making | 23 | 66 | -11 | 62 |
| Planning and Organization | 07 | 62 | 08 | 49 |
| Written Communications | 01 | 62 | -36 | 46 |
| Administrative Ability | 28 | 56 | 13 | 59 |
| Risk Taking | -28 | 53 | 31 | 51 |
| Resistance to Stress | 52 | 26 | 46 | 73 |

Note.—Decimals omitted.

TABLE 2

INTERCORRELATIONS^a AMONG ASSESSMENT SCALES AND CRITERIA

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | $r_{.6.5}$ |
|--|---|----|----|-----|----|----|------------|
| 1. Rating Scale 1—Activity | — | 43 | 37 | 27 | 49 | 78 | 72 |
| 2. Rating Scale 2—Administration | | — | 50 | 16 | 48 | 50 | 36 |
| 3. Rating Scale 3—Stress resistance | | | — | -15 | 26 | 25 | 15 |
| 4. Relative salary standing (external criterion) | | | | — | 10 | 37 | 37 |
| 5. Managers' potential evaluation (parallel criterion) | | | | | — | 46 | — |
| 6. Assessment program evaluation (internal criterion) | | | | | | — | — |

Note.—Decimals omitted.

^a $r = .29$ significant at $p < .05$ with $df = 45$.

program: (a) extent of active and aggressive participation in the interpersonal exercises, (b) ability as an independent decision maker and administrator, and (c) ability to function effectively in stressful situations and a willingness to take risks.

In the hierarchical factor analysis of the 25 traits included in the AT&T program, Bray and Grant (1966) identified 11 factors for their college graduate sample. Their analysis of higher order factors suggested that a major part of the variance was accounted for by several general factors reflecting overall program evaluations throughout most of their scales. However, considerable variance was accounted for by more specific factors. For the present study, the trait scales utilized for ratings appear more general than many of the AT&T rating scales and they reduce to two basic dimensions and possibly suggest a third. Based on these results, the trait ratings were collapsed into three summary scales and scores were developed for each individual: (a) Rating Scale 1—Activity: the mean of the six trait ratings loading most heavily on Factor 1. (b) Rating Scale 2—Administration: the mean of the four major trait ratings for Factor 2. (c) Rating Scale 3—Stress resistance: the mean of ratings on resistance to stress and risk taking.

Table 2 presents intercorrelations among these scales and the several "criteria" utilized in the research: the assessment program evaluation (internal criterion), the personnel jacket evaluation of potential (parallel criterion), and a current value criterion representing relative current salary standing, in thirds, of the individual in comparison with his peers (external criterion). Table 2 also

shows partial correlations between each scale and the internal criterion, controlling for the correlation with the parallel criterion as an indication of the contribution of criterion variance by each assessment program scale over and above variance associated with already available data in personnel records. Table 3 presents the correlations of the paper-and-pencil test scores—the two ability tests (Concept Mastery and SCAT—Numerical) and 20 scales derived from the battery of personality tests—and each of the three summary rating scales and the three criteria, as well as partial correlations of test scores with the internal criterion controlling for the parallel criterion.

As Table 3 suggests, there is some parallel between ratings obtained from the situational tests and the paper-and-pencil personality tests which lends a certain amount of credence to the constructs being measured. This is at least the case with regard to the construct of interpersonal activity as assessed by rating Scale 1 and by such personality test scales as GPP Ascendency, Background Survey—Self-Confidence, or SV Political which correlate significantly with the situational test ratings, suggesting that self-perceptions as measured with the personality tests and observer perceptions are to at least some extent parallel for this particular group. (It should be recognized that there was no contamination of the ratings from the personality test scores since these were not yet scored at the time the ratings were compiled.) There also appears to be some parallel between Scale 3 and such test scores as SDI—Occupational Level, Risk Taking, or SDI Initiative, also lending confidence to the constructs being evaluated. There

TABLE 3

CORRELATIONS* OF TEST SCALES WITH ASSESSMENT PROGRAM RATING SCALES AND WITH CRITERIA

| Test scale | Rating scale | | | Criterion measure | | | |
|---------------------------------------|--------------|------|------|-------------------------------------|---|--|-------------------------------|
| | 1. | 2. | 3. | 4. | 5. | 6. | |
| | 1 | 2 | 3 | Relative salary standing (external) | Manager's potential evaluation (parallel) | Assessment program evaluation (internal) | Assessment partial $r_{.6.5}$ |
| Concept Mastery (Total) | .06 | .04 | -.04 | .24 | -.02 | .10 | .12 |
| SCAT—Numerical | -.23 | -.03 | -.20 | .03 | -.13 | .03 | .10 |
| Study of Values—Theoretical | .04 | .25 | .08 | -.02 | .16 | .02 | -.06 |
| Study of Values—Economic | .17 | .26 | .18 | -.09 | .25 | -.01 | -.14 |
| Study of Values—Aesthetic | -.14 | -.01 | -.04 | -.04 | -.05 | -.14 | -.13 |
| Study of Values—Social | -.17 | -.17 | .08 | .00 | -.28 | -.23 | -.12 |
| Study of Values—Political | .38 | .25 | .17 | .16 | .33 | .25 | .12 |
| Study of Values—Religious | -.18 | -.06 | -.10 | .04 | -.05 | -.18 | -.18 |
| Background Survey—Management Key | -.17 | .10 | .12 | .05 | .22 | .09 | -.01 |
| Background Survey—Self-Confidence Key | .43 | .25 | .25 | -.05 | .31 | .27 | .15 |
| GPP—Ascendancy | .56 | .26 | .30 | .10 | .29 | .43 | .31 |
| GPP—Responsibility | -.34 | .01 | -.18 | .08 | -.26 | -.19 | -.08 |
| GPP—Emotional Stability | -.25 | -.09 | -.09 | -.06 | -.37 | -.32 | -.18 |
| GPP—Sociability | .26 | .28 | .21 | .01 | .37 | .23 | .07 |
| LOQ—Initiating Structure | .04 | .00 | .16 | .01 | .03 | -.13 | -.16 |
| LOQ—Consideration | -.15 | -.11 | .09 | .10 | -.13 | -.24 | -.20 |
| SDI—Intelligence | .24 | .20 | .07 | .14 | .09 | .26 | .25 |
| SDI—Supervisory Qualities | .29 | .12 | .07 | .26 | .06 | .32 | .33 |
| SDI—Initiative | .17 | .12 | .36 | -.10 | .08 | .10 | .07 |
| SDI—Self-Assurance | .25 | .32 | .18 | .15 | .05 | .33 | .35 |
| SDI—Occupational Level | .42 | .28 | .44 | .13 | .30 | .47 | .39 |
| Risk Taking | .36 | .26 | .36 | -.02 | .29 | .42 | .34 |

Note.—Decimals omitted.

* $r = .29$ significant at $p < .05$ with $df = 45$.

is less evident parallel between test data and situational data for Scale 2. The mental ability tests, Concept Mastery and SCAT (Quantitative), appear to be unique in this array of data and have their highest correlations with one another ($r = .31$).

While there was far from perfect agreement between the two management representatives who developed ratings of management potential from the 47 personnel jackets—their ratings correlated only .56—a mean of their individual predictions (the parallel criterion) does appear to cover much of the same ground that is covered by the 2-day program (the internal criterion), with one significant exception. As Table 2 shows, these two independent overall assessments of management potential correlate .46. Based on a comparison of the relationships between the rating scales and tests from the assessment program and each

of these overall evaluations, it is evident that they are very similar except for Rating Scale 1 dealing with interpersonal relationships; this scale correlates .78 with the overall program evaluation and .49 with the personnel jacket evaluation. This is hardly unexpected since the evaluators utilized only documents from the personnel file and did not interview and in most cases did not know personally the individuals they were evaluating. But for the other scales they seem to have overlapped fully as much predictor variance as the 2-day program.

DISCUSSION

This analysis suggests that there is probably not a great deal of intraindividual discrimination developed in the situational exercises utilized in the management assessment program described for this particular popula-

tion in this particular organization. There seems to be considerable overlap among the 12 trait ratings, and it appears that a limited number of concepts were actually evaluated.

Most of the scales derived from the several personality tests appear to make little unique contribution to the assessment evaluation (correlations with the internal criterion for 14 of the 20 scales are nonsignificant, and the largest is for the SDI Occupational Level as reported in Table 3). These results are probably representative of the very severe problems in the area of personality testing as discussed by Guion and Gottier (1965). While data for a few of the tests appear promising (e.g., GPP, SDI, Risk taking), from multiple regression analyses it is not clear that they provide much incremental variance over situational measures in explaining the criteria utilized in this study. The patterns of overlap between these test scales and the situational rating data evident in Table 3 suggests a similar interpretation. This, of course, does not address the issue of the utility of these personality test scales in a long-term predictive validation.

The major component of assessment—both from situational and from personality tests—seems to be an evaluation of interpersonal behavior. Mental ability measurements contribute essentially nothing to this prediction of managerial success.

The data suggest that traditional approaches to the assessment of management potential in the form of a careful evaluation of personnel records and employment history (our parallel criteria) can perhaps provide much of the same information which evolves from the lengthy and expensive 2-day assessment program, as many of the evaluation components which emerge from the program correlate as highly with this rating as they do with the internal criterion. However, the partial correlations indicate that these are not completely equivalent evaluations of managerial potential. The biggest discrepancy is for Rating Scale 1—the evaluation of interpersonal behavior. Quite possibly, if this traditional assessment were accompanied by an extensive personal interview, as would be the case in real life, traditional approaches might

be found to be even more effective in comparison with the situational exercises. The results suggest that such a systematic evaluation of past history might fruitfully be included in overall assessments of management potential compiled in programs such as this.

This comparison raises an issue which can easily be overlooked in the rush to institute a management assessment program: a clear definition of just what it is hoped the program will do and where in an individual's career it can be most fruitfully utilized. If such a program is looked to for evaluating candidates for middle management or higher positions, then one may justifiably question the appropriateness of two days of situational exercises for evaluating management potential as opposed to a careful review of prior job history and accomplishment. On the other hand, if the focus is on the *early* identification of potential where little job history has accrued, then the assessment center is probably a very effective means of synthesizing a rather close approximation to the type of potential prediction which would eventually evolve through on-the-job performance. And, as hinted at the outset, there are numerous other potential benefits which can be derived from such a program among relatively young premanagement candidates.

The analysis still does not adequately talk to the question of validity, and the ultimate test of the utility of such an approach would have to be based on a careful predictive validation strategy—an approach which all too often is essentially impossible considering the time span of the prediction involved and the pressures in real-life organizations for immediate evaluation of programs and implementation of new techniques. However, the lack of concurrent validity in this analysis argues strongly for additional research to assess the predictive validity of this particular management assessment methodology in this particular setting. While in all probability our external criterion of relative salary standing was not as reliable a criteria of current worth as one might wish, there was also no correlation between program ratings and other concurrent criteria of performance ratings or normalized rankings.

This is not necessarily an indictment of the program's possible value and perhaps should be viewed positively since the program's intent was predictive rather than concurrent validity, and the rating which was evolved was of a future criterion variable. In this situation, a fair validation would require allowing adequate time for meaningful criterion variance to develop. The AT&T data and other unpublished results suggest that predictive validities improve as the time interval between data collection and criterion measurement approximate the time span for the predictive assessment (Bray & Grant, 1966, p. 18).²

At the very least, the lack of demonstrated concurrent validity in this study calls for continued validation research to ensure proper utilization of this type of information which potentially can have such far-reaching implications for individual careers. The present analysis does suggest, however, that the assessment center evaluation contains reliable variance which may be associated with management potential. And the research suggests some areas for improvement of the program within this particular environment.

² In this regard, it is interesting that the personality test scales in this study which are most strongly correlated with the assessment evaluation of managerial potential clearly have the flavor of upward mobility—occupational level, ascendancy, risk taking, self-assurance, or supervisory qualities.

REFERENCES

- ALBRECHT, P. A., GLASER, E. M., & MARKS, J. Validation of a multiple-assessment procedure for managerial personnel. *Journal of Applied Psychology*, 1964, **48**, 351-360.
- BRAY, D. W., & CAMPBELL, R. J. Selection of salesmen by means of an assessment center. *Journal of Applied Psychology*, 1968, **52**, 36-41.
- BRAY, D. W., & GRANT, D. L. The assessment center in the measurement of potential for business management. *Psychological Monographs*, 1966, **80**(17, Whole No. 625).
- CARROLL, J. B. IBM 709/7090 program for generalized analytic solution in factor analysis. Harvard University, 1960. (mimeo)
- FLANAGAN, J. C. Some considerations in the development of situational tests. *Personnel Psychology*, 1954, **7**, 461-464.
- GHISELLI, E. E. The validity of management traits in relation to occupational level. *Personnel Psychology*, 1963, **16**, 109-114.
- GHISELLI, E. E. The validity of a personnel interview. *Personnel Psychology*, 1966, **19**, 389-394.
- GRANT, D. L., KATKOVSKY, W., & BRAY, D. W. Contributions of projective techniques to assessment of management potential. *Journal of Applied Psychology*, 1967, **51**, 226-232.
- GREENWOOD, J. M., & McNAMARA, W. J. Interrater reliability in situational tests. *Journal of Applied Psychology*, 1967, **51**, 101-106.
- GUION, R. M., & GOTTIER, R. F. Validity of personality measures in personnel selection. *Personnel Psychology*, 1965, **18**, 135-164.
- HARDESTY, D. L., & JONES, W. S. Characteristics of judged high-potential management personnel—the operations of an industrial assessment center. *Personnel Psychology*, 1968, **21**, 85-98.
- LAURENT, H. A summary of early identification of management potential. Research Project, Social Science Research Division of the Employee Relations Department, Standard Oil Co., 1961.
- PRIEN, E. P. Assessments of higher-level personnel: V. An analysis of interviewers' predictions of job performance. *Personnel Psychology*, 1962, **15**, 319-334.
- WILLIAMS, L. K. Some correlates of risk taking. *Personnel Psychology*, 1965, **18**, 297-310.

(Received October 30, 1968)

PROCESSING AUDITORY INFORMATION: INTERFERENCE FROM AN IRRELEVANT CUE

J. RICHARD SIMON¹ AND A. M. SMALL, JR.²

University of Iowa

In a choice RT task, 64 Ss pressed either a right- or left-hand key in response to directional commands provided by 400 and 1000 cps tones. On monaural trials, RT was significantly faster when the meaning of the tonal command corresponded with the ear in which it was heard (corresponding trials) than when it did not (noncorresponding trials). A comparison of monaural with binaural RT indicated that this Tonal Command \times Ear Stimulated interaction was due to interference on the noncorresponding monaural trials rather than facilitation on the corresponding trials.

In an experiment concerned with reaction time (RT) to monaurally presented verbal commands, Simon and Rudell (1967) discovered an extremely potent phenomenon, namely, that the speed of processing the symbolic content of a command was affected by the ear in which the command was heard. Their Ss responded significantly faster when the content of the command corresponded to the ear stimulated (i.e., "right" in right ear or "left" in left ear) than when it did not (i.e., "right" in left ear or "left" in right ear). Results clearly suggested that the auditory display provided two cues, one relevant (content of command) and the other irrelevant (ear stimulated), and that the time required to process the former was somehow affected by the presence of the latter. Left unclear was the basic nature of the cue provided by the ear stimulated. Did the cue facilitate information processing on trials where it corresponded with the symbolic content of the command or did it interfere with information processing on trials where it did not correspond with the content of the command? Could the cue have operated *both* to facilitate responding on the corresponding trials *and* to interfere with responding on the noncorresponding trials? In other words, how was the Command \times Ear Stimulated interaction produced? The primary purpose of the present experiment was to answer these ques-

tions. A second purpose of this study was to determine the generality of the Command \times Ear Stimulated interaction. Is the phenomenon limited to situations involving verbal commands or does it also occur when simple stimuli such as pure tones are used to provide the relevant directional information?

METHOD

Apparatus. The apparatus provided a measure of choice RT to a series of tones presented to S through Telephonics TDH-39, 300 ohm earphones. The earphones were mounted in NAF-48490-1 cushions and fixed to a standard headband. The Ss' task was to press the correct one of two finger keys as soon as possible after hearing the tone. A Hunter klockounter started when the tone was presented and stopped when S pressed a key. Depressing the key also signaled E as to which key (or keys) had been pressed. The keys, which S operated with his right and left index fingers, were mounted 12 in. apart on a table in front of him. Two Hewlett-Packard Model 200AB audio oscillators were used to generate tones of 400 and 1000 cps. Silent switches permitted E to present one tone or the other to either the left ear, the right ear, or to both ears simultaneously. On monaural trials, the output SPL was 99 db. On binaural trials, the output SPL was reduced to 93 db. so as to yield approximately the same loudness as on the monaural trials (Caussé & Chavasse, 1942). A warning light was presented 2 sec. prior to the onset of each tone, and there was a 7-sec. interval between trials.

Subjects. The Ss were 32 male and 32 female University of Iowa undergraduate volunteers. All Ss reported having normal hearing.

Procedure and experimental design. Each S performed on two blocks of trials, one block involving monaural stimulation and the other involving binaural stimulation. On the monaural trials, either a high-pitched tone (1000 cps) or a low-pitched tone (400 cps) was presented to one ear. The Ss had no way of knowing, prior to the presentation of a tone,

¹ Requests for reprints should be sent to J. Richard Simon, Department of Psychology, University of Iowa, Iowa City, Iowa 52240.

² The authors acknowledge the assistance of Michael Thall in collecting the data.

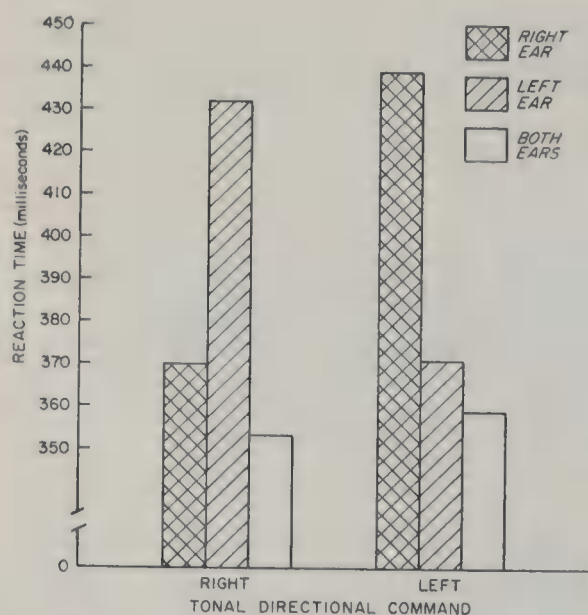


FIG. 1. Reaction time to tonal directional commands as a function of ear(s) stimulated.

which ear would be stimulated or what the tone would be. Half of *Ss* were instructed to press the right key when they heard the high-pitched tone and to press the left key when they heard the low-pitched tone. The other half of *Ss* were given the opposite tone-key rule. There were 56 monaural test trials in which the 400 and 1000 cps tones were presented equally often to each ear in a predetermined random sequence. The test trials were preceded by eight practice trials in which each tone was presented twice to each ear in a random sequence. On the binaural trials, either the 400 or the 1000 cps tone was presented to both ears simultaneously. Eight practice trials and 56 test trials were given in the same random sequence used in the monaural block. Each *S* performed both monaural and binaural blocks using the particular tone-key rule to which he was originally assigned.

Half of the males and half of the females performed the monaural trials first while the other half performed the binaural trials first. Each sex \times sequence group was further subdivided by assigning eight *Ss* to one tone-key rule and eight to the opposite rule. Finally, in order to balance out any differences which may have existed between stimulus channels, the earphones were reversed for half of *Ss* in each subgroup.

RESULTS

Median RTs were computed for each *S* for each of the six treatment conditions, that is, right and left tonal commands in the right ear, the left ear, and in both ears simultaneously. An analysis of variance of the monaural trials revealed no differences as a function of ear stimulated, tonal command, sex, or order.

The only significant effect was the expected tonal command \times ear stimulated interaction, $F(1, 60) = 276.35$, $p < .001$. Figure 1 shows that RT was markedly faster when the right command was heard in the right ear than when it was heard in the left ear (370 vs. 432 msec.). Similarly, RT to the left command was faster when it was heard in the left ear than when it was heard in the right (371 vs. 439 msec.). Clearly, *Ss* responded significantly faster on trials where the symbolic content of the command corresponded with ear stimulated (corresponding trials) than on trials where it did not (noncorresponding).

Since the major purpose of the experiment was to determine whether the tonal command \times ear stimulated interaction reflected a facilitation of information processing on the corresponding trials or an interference with information processing on the noncorresponding trials, additional comparisons of binaural with monaural RT were conducted. Binaural RT (356 msec.) was significantly faster than average RT on the noncorresponding monaural trials (435 msec.)— $F(1, 60) = 269.97$, $p < .001$. Binaural RT was also significantly faster than average RT on the corresponding monaural trials (371 msec.)— $F(1, 60) = 11.82$, $p < .01$. Right responses tended to be faster than left responses, but this difference reached significance ($p < .05$) in only one of the analyses.

DISCUSSION

Results of this experiment clearly indicated that the tonal command \times ear stimulated interaction on the monaural trials was a result of interference on the noncorresponding trials rather than facilitation on the corresponding trials. This conclusion was reached after comparing binaural RT with RT on both noncorresponding and corresponding monaural trials in turn. On the monaural trials, the ear in which the tone was heard provided an irrelevant cue which, of course, was absent on the binaural trials. Thus, binaural RT provided an appropriate baseline for evaluating the effect of the irrelevant cue. On the noncorresponding monaural trials, the irrelevant cue (ear stimulated) apparently conflicted with processing the relevant cue

(frequency of the tone) resulting in slower RT than on the binaural trials. On the corresponding monaural trials, the irrelevant cue coincided with the relevant cue (i.e., right command in right ear and left command in left ear), but this correspondence did not facilitate information processing. In fact, RT on the corresponding monaural trials was also significantly slower than binaural RT. The reason for this latter finding is not clear. Perhaps, the corresponding monaural trials were slowed by the presence of the noncorresponding trials in the same block. Alternatively, it may be that binaural stimulation per se results in faster RT than monaural stimulation.

Another important outcome of this experiment was the demonstration that the command \times ear stimulated interaction, heretofore only observed with verbal directional commands (Simon & Rudell, 1967; Simon, 1968), also occurred when pure tones were used to signal the appropriate response. Thus, it appears that the interaction reflects a basic and general phenomenon which exists inde-

pendently of whether the command is communicated verbally or nonverbally. It also appears that the interaction is unrelated to prior symbolic associations since, in contrast to verbal directional commands, the tones had no implicit directional significance. While much remains to be learned about the exact nature of the interference phenomenon, results to date clearly underscore its potency and emphasize the importance of hitherto unrecognized spatial cues in decoding auditory displays.

REFERENCES

- CAUSSÉ, R., & CHAVASSE, P. Différence entre l'écoute binauriculaire et monauriculaire pour la perception des intensités supraliminales. *Comptes Rendus de la Société de Biologie*, Paris, 1942, **136**, 405-406.
- SIMON, J. R., & RUDELL, A. P. Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, 1967, **51**, 300-304.
- SIMON, J. R. Effect of ear stimulated on reaction time and movement time. *Journal of Experimental Psychology*, 1968, **78**, 344-346.

(Received November 7, 1968)

IMAGE OF INDUSTRIAL PSYCHOLOGY AMONG PERSONNEL ADMINISTRATORS

GEORGE C. THORNTON III

Colorado State University

A survey was conducted of impressions of industrial psychologists among a national sample of personnel administrators. Results showed 11% of all companies and 20% of the largest ones employ a psychologist full time, and 25% employ one or more as a consultant. Fifty percent of the respondents felt it would be desirable to have a psychologist in the company and 75% felt he would increase productivity and satisfaction. Ratings of perceived past contributions, future usefulness, and need for further research in 12 areas of specialization are presented. Comparisons are made with previous surveys over a 20-yr. period.

While it has been noted generally that psychologists working in industry have been utilized more frequently and more broadly over the past decades, there is little systematic evidence of these trends. This paper reports a survey of the impressions of a large sample of personnel administrators representative of all American industries with regard to their impressions of the current use and potential contributions of psychology in industry. Comparisons are made with previous surveys (Feinberg & Lefkowitz, 1962; Stagner, 1946; Tiffin & Prevratil, 1956).

METHOD

Questionnaire. A questionnaire was constructed to obtain the following information: demographic data about the respondent (age, amount and type of education, and authority to hire an industrial psychologist), descriptive data about the company (size, industry, amount of unionization, size of personnel department, and whether it employed an industrial psychologist), and the respondent's impressions of industrial psychology (general rating of contribution toward increased productivity and employee satisfaction, and specific rating of past contribution, future contribution, and need for future research in 12 areas of specialization).

Subjects. The Ss were a 20% random sample of personnel administrators chosen from the 1966 directory of the American Society for Personnel Administration. Of the 600 questionnaires mailed, 319 usable ones were returned. The companies employing the respondents and nonrespondents did not differ significantly in size and region of the country.

RESULTS

The sample was typically in their 30's and 40's, had a bachelor's degree (54%), were trained in business administration or industrial relations (43%), and had the authority to employ a psychologist (37%). Only 12% of the sample of Feinberg and Lefkowitz (1962) reported such authority. The employing companies consisted of approximately equal numbers of small (less than 1,000 employees), medium (1,000-4,999), and large (over 5,000) companies having a wide range of union representation and sizes of personnel departments.

Approximately half (46%) of the companies employ an industrial psychologist full time or on a consulting basis. Tiffin and Prevratil (1956) found the comparable figure to be 28.8%, the difference being significant at the .01 level. Eleven percent of all companies and 20% of the largest companies employed an industrial psychologist full time. Stagner (1946) in his survey of large corporations found that 30% employed a professionally trained psychologist full time.

The questions and responses concerning the impressions of personnel administrators toward psychologists are contained in Table 1. One-half of the sample felt it would be desirable to have an industrial psychologist actually in the company, and about three-fourths felt such a person could be useful in increasing both worker productivity and satisfaction. In considering the specific areas where psychologists have been of value in the past, Ss most

¹ Requests for reprints should be sent to the author, Department of Psychology, Colorado State University, Fort Collins, Colorado 80521.

TABLE 1

RATINGS OF DESIRABILITY, USEFULNESS, AND
NEEDED RESEARCH OF INDUSTRIAL
PSYCHOLOGISTS
(*N* = 319)

| | | Yes | No |
|---|------------|----------------------|-----------------|
| Would (Do) you consider it desirable to have a professionally trained industrial psychologist in your company? | | 53% | 43% |
| Do you think the services of an industrial psychologist could be useful in your company in increasing (a) productivity? | | 72% | 22% |
| (b) satisfaction? | | 76% | 18% |
| Subareas | Past value | Usefulness in future | Needed research |
| Employee selection | 42% | 46% | 31% |
| Employee training | 20% | 32% | 20% |
| Managerial selection | 47% | 59% | 39% |
| Managerial training | 32% | 59% | 39% |
| Performance appraisal | 20% | 38% | 38% |
| Job evaluation | 10% | 14% | 8% |
| Labor relations | 14% | 15% | 14% |
| Employee motivation/attitude (morale) surveys | 24% | 60% | 47% |
| Safety and accident prevention | 4% | 11% | 13% |
| Organization analysis and planning | 17% | 25% | 24% |
| Human factors engineering | 11% | 34% | 25% |
| Consumer behavior | 2% | 8% | 9% |
| Other | 17% | 3% | 5% |

frequently checked employee and managerial selection (42% & 47%), employee and managerial training (20% & 32%), motivation and morale surveys (24%), and performance appraisal (20%). In the areas of safety, labor relations, and consumer behavior, few of the sample saw any contribution. In terms of usefulness in the future, the same general pattern is noted. In the areas of managerial training, performance appraisal, employee motivation, and human factors engineering, there seems to be the feeling that more could be done in the future than has been done in the past. The evaluations of potential usefulness in the specific areas by this group is markedly similar to those of the previous samples. One area, organizational analysis and planning, was included in the present study and not in previous studies. Twenty-five percent felt psychologists could be of service here, in comparison with Feinberg and Lefkowitz's (1962) conclusion: "The executives in our sample never hired a psychologist to deal with broad categories such as research or organizational structure [p. 110]."

In terms of needed research, the areas of motivation and attitude surveys were rated highest (47%). A number of other areas were checked by 25–35% of the group. The areas of job evaluation, safety, labor relations, and consumer behavior were checked very infrequently.

DISCUSSION

It is felt the sample was a representative, appropriate, and influential one, and thus allows meaningful statements about the current impressions of personnel administrators toward psychologists in industry.

While it does not appear that the percentage of larger companies employing a full-time psychologist has increased over the past several years, it may be that larger numbers of smaller one are, and it seems that more companies of all sizes are using psychological consultants. When executives have been asked over the past 20 yr. if they consider it desirable to have a professionally trained psychologist in their company, an affirmative answer has been given by virtually the same percent of respondents: 1946—53%; 1956—54.5%; 1962—66%; and 1967—53%. Even though 50% think it would be desirable to employ an industrial psychologist, only 11% do so. This would indicate there is a great opportunity for qualified and interested persons. Stagner (1946) found the same gap 20 years ago and came to the same conclusion.

The large majority of the personnel administrators felt that the services of a psychologist are useful. It was somewhat surprising that the evaluations of the specific areas were so similar to the evaluations in previous studies. Industrial psychologists have made advancements in understanding in many areas, but it seems these have gone unrecognized or have not been successfully applied to personnel problems. It should not be inferred that the need for research has diminished; the results of this study alone would argue that personnel administrators are looking for advances in a number of different areas through additional research.

REFERENCES

- FEINBERG, M. L., & LEFKOWITZ, J. Image of industrial psychology among corporate executives. *American Psychologist*, 1962, 17, 109-111.
- STAGNER, R. Attitudes of corporate executives regarding psychological methods in personnel work. *American Psychologist*, 1946, 1, 540-541.
- TIFFIN, J., & PREVRATIL, W. Industrial psychology in the aircraft industry. *American Psychologist*, 1956, 11, 246-248.

(Received November 14, 1968)

(Continued from page 398)

- Feedback Factors in Steering and Tracking Behavior: Karl U. Smith* and Vernon Putz: Department of Psychology, University of Wisconsin, Charter at Johnson, Madison, Wisconsin 53706.
- Influence of Group and Situational Differences on the Applicability of a Personnel Test: I. van W. Raubenheimer*: Department of Industrial Psychology, University of South Africa, P. O. Box 392, Pretoria, South Africa.
- Effect of Applicant Sample on Ratings of Valid Information in an Employment Setting: Robert E. Carlson*: Research Division, Life Insurance Agency, Management Association, 170 Sigourney Street, Hartford, Connecticut 06105.
- Effects of Piece Rate Overpayment on Productivity: Ian Wood and Edward E. Lawler, III*: Department of Administrative Sciences, Yale University, 2 Hillhouse Avenue, New Haven, Connecticut 06520.
- Effects of Sound on Postural Control: Karl U. Smith,* Gerhard Lingh, and Roger Eischens: Department of Psychology, University of Wisconsin, Charter at Johnson, Madison, Wisconsin 53706.
- Intrasubject Prediction of Preferences for Organization Types: James L. Sheard*: School of Systems and Logistics, Wright-Patterson Air Force Base, Ohio 34322.
- Validation of Employment Tests for Telephone Company Installation and Repair Occupations: Donald L. Grant* and Douglas W. Bray: Personnel Manager, Research, American Telephone and Telegraph Company, Room 2122B, 195 Broadway, New York, New York 10007.
- Cue Conflict Related to Perceptual Style: Gerald V. Barrett,* Carl L. Thornton, and Patrick A. Cabe: Management Research Center, University of Rochester, 1409 Temple Building, Rochester, New York 14604.
- Task Difficulty as a Moderator of Long-Range Prediction: W. E. Dodd,* H. B. Wollowick, and W. J. McNamara: Executive Resources Department, IBM Corporation, Armonk, New York 10504.
- The Guilty Knowledge Technique: Application of Signal Detection Measures: Gerson Ben Shakhar* and Sol Kugelmass: Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel.

* Asterisk indicates author for whom address is supplied.

Journal of applied psychology

Table of Contents

| | |
|---|-----|
| Prediction of Job Performance from Assessment Reports: Use of a Modified Q-Sort Technique to Expand Predictor and Criterion Variance. <i>Garland Y. DeNelsky and Michael G. McKee</i> | 439 |
| Psychodynamics of Intrinsic and Extrinsic Job Orientation. <i>S. D. Saleh and T. G. Grygier</i> | 446 |
| Comparison of Several Patterns of Communication. <i>Charles J. O'Leary and Morton Goldman</i> | 451 |
| Work Values and Job Satisfaction. <i>Milton R. Blood</i> | 456 |
| Use of Leadership Powers in Industry. <i>David Kipnis and Joseph Cosentino</i> | 460 |
| Impact of Employee Participatiton in the Development of Pay Incentive Plans: A Field Experiment. <i>Edward E. Lawler III and J. Richard Hackman</i> | 467 |
| Consumer Attitudes toward Auto versus Public Transport Alternatives. <i>Frank T. Paine, Allan N. Nash, Stanley J. Hille, and G. Allen Brunner</i> | 472 |
| Correlates of Employee Evaluations of Pay Increases. <i>J. R. Hinrichs</i> | 481 |
| Effects of Performance on Leadership, Cohesivoness, Influence, Satisfaction, and Subsequent Performance. <i>George P. Farris and Francis G. Lim, Jr.</i> | 490 |
| Skimming Lists of Food Ingredients Printed in Different Brightness Contrasts. <i>E. C. Poulton</i> | 498 |
| Effect of Interviews on Teacher Selection Decisions. <i>Dale L. Bolton and Michael E. Hickey</i> | 501 |
| On the Use of Alternative Organisms. <i>Douglas A. Bernstein and Thomas M. Alloway</i> | 506 |
| A Simplified Method for Increasing Mail-Sorting Efficiency. <i>Thomas Rex Long, Dell Lebo, and Dorothy Ogden</i> | 510 |
| A Situational Approach to Leadership Effectiveness. <i>Walter Hill</i> | 513 |
| Relevance of Rater-Ratee Acquaintance in the Validity and Reliability of Ratings. <i>Norman E. Freeberg</i> | 518 |
| Effects of Organizational Structure on Correlations between Member Abilities and Group Productivity. <i>Gordon E. O'Brien and A. G. Owens</i> | 525 |
| Attitudes of Professional Appraisers toward Aesthetics and Appraisal Activities. <i>Jack M. Wright and J. H. Lemly</i> | 531 |
| Distortion of Drivers' Estimates of Automobile Speed as a Function of Speed Adaptation. <i>Frank Schmidt and Joseph Tiffin</i> | 536 |
| List of Manuscripts Accepted. | 512 |

This is the last issue of Volume 53.
Volume Title Page and Contents appear herein.

Editor

KENNETH E. CLARK
University of Rochester

Consulting Editors

GEORGE E. BRIGGS
Ohio State University

MARVIN D. DUNNETTE
University of Minnesota

NORMAN FREDERIKSEN
Educational Testing Services

LEONARD D. GOODSTEIN
University of Cincinnati

EDWIN R. HENRY
University of Rochester

JOHN HOLLAND
American College Testing Program

CLIFFORD E. JURGENSEN
Minneapolis Gas Company

DONALD G. MARQUIS
Massachusetts Institute of Technology

LAURENCE S. MCGAUGHERAN
University of Houston

QUINN MCNEMAR
University of Texas

WARREN T. NORMAN
University of Michigan

HAROLD F. ROTHE
Beloit Corporation

THOMAS A. RYAN
Cornell University

ALEXANDER C. WEXMAN
Psychological Corporation

CLARK L. WILSON
Instituto Centroamericano de Administracion de Empresas

This journal gives primary consideration to original investigations in any field of applied psychology except clinical psychology, although a descriptive or theoretical article may be accepted if it represents a special contribution in an applied field. Quantitative investigations of interest or value to psychologists working in the following broad fields will be considered: personnel research in business, industry, and government; engineering psychology; industrial working conditions; research on opinion and morale factors; job analysis and classification research; market and advertising research; vocational and educational prognosis, diagnosis, and guidance at the secondary and college level.

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or they may be obtained from the editor or from the APA Central Office.

Manuscripts should be addressed to the editor-elect:

Dr. Edwin A. Fleishman, Director
American Institutes for Research
8555 Sixteenth Street
Silver Spring, Maryland 20910

Authors are strongly advised to follow the general directions given in the *Publication Manual of the American Psychological Association* (1967 Revision). Special attention should be given to the section on the preparation of the references (p. 39). In reference lists, give journal titles in full; do not abbreviate. *All copy must be double spaced, including the references.* All manuscripts should be submitted in *duplicate*, one of which should be an original typed copy; author's name should appear only on title page. Dittoed and mimeographed copies are not acceptable and will not be considered. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail and to check carefully the typing of the final copy.

The following policies govern reprinting or other forms of reproduction of APA-copyrighted journal materials. Permission by the American Psychological Association must be requested for reprinting all tables, figures, and text of more than 500 words (total from one article) in length. There will be a charge of \$5.00 per page for such permission except when the requester is the author of the article to be reprinted, or the article is to be reproduced in a limited number of copies solely for instructional purposes.

HELEN ORR
Managing Editor

CLAIRE PILLIOD
Technical Editor

HENRY S. SALMON
Advertising and
Business Manager

ROBERT D. GORMAN
Subscription Manager

Subscriptions, orders, and business communications should be addressed to the American Psychological Association, 1200 Seventeenth St., N. W., Washington, D. C. 20036. Address changes must reach the subscription office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Second-class postage paid at Lancaster, Pa.

Published bimonthly by the
American Psychological Association
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W.
Washington, D. C. 20036

\$10.00 per volume

\$2.00 per issue

PREDICTION OF JOB PERFORMANCE FROM ASSESSMENT REPORTS: USE OF A MODIFIED Q-SORT TECHNIQUE TO EXPAND PREDICTOR AND CRITERION VARIANCE ¹

GARLAND Y. DENELSKY ² AND MICHAEL G. MCKEE ³

Central Intelligence Agency

Predictions of performance and personality characteristics made on the basis of preemployment psychological assessment reports were compared with subsequent performance evaluations contained in the fitness reports of 32 government employees. Seven psychologists reviewed the assessment reports as a basis for predicting overall job effectiveness and specific performance and personality characteristics. They then reviewed the narrative section of each individual's fitness report as a basis for rating the overall effectiveness of each person. Ratings were made using a modified Q-sort technique that reliably expanded the variances of predictor and criterion variables. A significant positive relationship was found between predicted and actual effectiveness. In addition, the psychologists were able to predict specific performance and personality dimensions on a significantly better than base-rate basis.

Over the past 20 years, with the 1948 Office of Strategic Services volume, *Assessment of Men*, lighting the way, there has been a steady if slow flow of research on the predictive validity of clinical assessment, using multiple methods for obtaining information about individuals. Taft (1959) provides a comprehensive review of the earlier studies. Studies by Bray and Grant (1966), Hilton, Bolin, Parker, Taylor, and Walker (1955), Campbell, Otis, Liske, and Prien (1962), Trankell (1959), Dicken and Black (1965), and Albrecht, Glaser, and Marks (1964) report significant positive correlations between assessment predictions and performance criteria. The results of some studies, however, have

cast doubt upon the predictive efficacy of assessment procedures (Holtzman & Sells, 1954; Kelly & Fiske, 1951).

Bray and Grant (1966) summarized the research to date as follows:

Though no firm conclusions regarding the predictive validities of multiple assessment procedures can be drawn from the rather mixed findings of published research, it does appear clear that the more accurate predictions were obtained where the performance to be predicted was clearly defined, the assessment results did not restrict the range of subsequent criterion performance, and the criterion measures employed were not limited by low reliability and questionable validity [p. 2].

Unfortunately, it is usually impossible to meet the above conditions in applied assessment; the job duties are heterogeneous and ill defined; criterion performance is restricted in range by selection on the basis of assessment results; the criterion measure is based on standard organizational evaluation reports and, as such, is of questionable validity. A variety of raters and a variety of jobs, with

¹ The views expressed in this article are those of the authors and do not necessarily reflect an official position of the Central Intelligence Agency.

² Requests for reprints should be sent to Garland Y. DeNelsky, Central Intelligence Agency, Washington, D. C. 20505.

³ Now at the Cleveland Clinic, Cleveland, Ohio.

the clearly inept performers screened out, tend to lower the correlations of predictors and rated job performance. Many elements in a study of assessment *au naturel* coalesce to lower validity, and the question is whether assessment has value within these limitations and whether it can predict performance in an ongoing occupational setting.

The purpose of the present study was to determine if predictive validity can be demonstrated for psychological assessments within a natural setting when a special rating technique that increases predictor and criterion variability is used. The specific focus of investigation was the assessment report; the major question was whether preemployment psychological assessment reports do predict the subsequent performance of those individuals who are hired.

METHOD

Subjects

Fitness reports (routine performance evaluations about one-half page in length) were obtained on 32 male employees who had been working overseas for 1 yr. or more. Assessment reports were available on all 32. These individuals had been assessed 12–57 mo. earlier by one of eight psychologists; the median interval between assessment and fitness reports was 20 mo. The original assessments varied slightly from case to case but typically included intellectual, personality, attitudinal, and interest testing in addition to one or more depth interviews. The assessment reports were typically one or two pages long and contained descriptions of the individual's strengths and weaknesses as well as a summary recommendation.

All 32 men were overseas at the time their fitness reports were prepared. Although it was not possible to determine how many different supervisors had actually been responsible for this group, it was established that none of the field supervisors had seen their assessment reports. The total of 32 men was divided into two groups. Each of these groups (which will be referred to as Group 1 and Group 2) contained 16 men. The two groups were judged separately; in fact, several months intervened between the judging of Group 1 and Group 2.

Seven staff psychologists served as judges. All had experience in assessing overseas candidates.

Procedure

Trait prediction. In the first phase of the study for both groups, each of the judges was given the 16 original assessment reports, together with a specially designed Trait Rating Sheet for each *S*. The Trait Rating Sheet listed 25 performance and personality

traits that had been abstracted from the narrative sections of the total group of fitness reports of the employees in the study. Performance ratings included such dimensions as response to supervision, accuracy of work, speed of learning, and supervisory effectiveness; personality ratings included such dimensions as judgment, maturity, flexibility, and self-confidence. Approximately half of the 25 dimensions could be described as personality variables; the other half pertained to job performance. The judges were instructed to form an impression of each of the men from the assessment report, and, on the basis of this impression, to predict whether each individual would be discussed favorably or unfavorably on each trait in his fitness report (assuming, of course, that he would be discussed on all dimensions—a slightly unrealistic situation since no employee was mentioned on more than 12 of the 25 dimensions). For those individuals mentioned favorably or unfavorably on a given dimension in their fitness reports, it was possible to determine if the predictions made by psychologists were in the same direction as the actual descriptions of the individuals in their fitness reports.

Q sorts of assessment and fitness reports. Following his completion of the Trait Rating Scales, each judge was asked to sort the assessment reports of the 16 men of each group into five categories corresponding to his prediction of each individual's overall effectiveness in a typical overseas work situation of the type to which these men were assigned. In order to eliminate variance due to differing frames of reference on the part of the seven judges, a modified *Q*-sort distribution was used; assessment reports were to be assigned to five categories, ranging from a predicted worst performance to a predicted best performance with 1, 4, 6, 4, and 1 individuals assigned to the respective categories. Score values of 1, 2, 3, 4, and 5 (best) were assigned to the five categories.

Following the *Q* sort of assessment reports on the basis of predicted overall effectiveness, each judge was assigned the task of *Q* sorting, in the same manner as before, each group of 16 individuals on the basis of actual overall effectiveness as described in narrative form in their fitness reports. The names of the 16 men were deleted from the fitness reports; thus the judges had no way of knowing which of the assessment reports and fitness reports had been written for the same persons.

It should be noted that the prediction situation as structured in this study was different from the usual design of studies with similar objectives. Instead of being given test scores and other psychometric and background data and being required to weight this "raw" information in order to make predictions of future behavior, the judges in this study were asked to formulate predictions on the basis of finished assessment reports. Thus, the judges in the present study were placed in a role similar to the consumer of psychological assessment reports: They were to make predictions on the basis of someone else's analysis and interpretation of first-hand data. Dicken and Black (1965) used a similar method,

TABLE 1
ANALYSIS OF VARIANCE RELIABILITY COEFFICIENTS
FOR ASSESSMENT- AND FITNESS-REPORT RATINGS

| Rating | Coefficient for single rating | | Coefficient for composite rating | |
|-------------------|----------------------------------|---------|-------------------------------------|---------|
| | Group 1 | Group 2 | Group 1 | Group 2 |
| Assessment report | .63 | .66 | .92 | .93 |
| Fitness report | .59 | .74 | .91 | .95 |

commenting that “the ratings are thus two interpretive steps removed from the original test data [p. 36].”

RESULTS

Prediction of Overall Effectiveness

Before relating assessment-report predictions to fitness-report ratings, it was necessary to establish the reliability of the judgments made by the judges on both measures.

Table 1 presents the analysis of variance reliability coefficients for the assessment- and fitness-report judgments. It is evident from this table that the reliabilities, particularly of the average or composite ratings for each individual by all judges, are quite satisfactory. Despite several judges’ comments that the task of making the ratings was a difficult one, there was substantial agreement among judges on both the assessment-report and the fitness-report ratings.

The answer to the primary question of this study—whether judges can predict, on the basis of psychological assessment reports, performance in actual field situations as judged from fitness-report narratives 12–57 mo. later—can be approached from a number of directions. Perhaps the single most meaningful approach is to correlate the composite assessment-report predictions of the seven judges for each of the 16 individuals in each group with the composite judged effectiveness of the same individuals based on fitness reports. The resulting correlations, presented in Table 2, indicate that with the total sample of 32 men, there is a significant positive relationship between the overall or composite predictions of effectiveness based on assessment reports and actual effectiveness as judged from fitness reports.

TABLE 2
CORRELATIONS BETWEEN COMPOSITE ASSESSMENT-
REPORT PREDICTIONS AND FITNESS-
REPORT EVALUATIONS

| Group | N | r |
|------------------|----|------|
| 1 | 16 | .42 |
| 2 | 16 | .25 |
| 1 and 2 combined | 32 | .32* |

* $p < .05$, one-tailed test.

Another way of illustrating the relationship between assessment and fitness reports is shown in Table 3. Of those 17 men with average or above assessment ratings, 12 (71%) received average or above fitness ratings, while only 6 (40%) of the 15 men with below-average assessment ratings received average or above fitness ratings.

Table 4 presents correlations between the individual judge’s assessment ratings and the composite fitness ratings (for Groups 1 and 2 combined). Assuming the composite of the fitness-report ratings by all judges is the best single measure of actual performance, the psychologists varied in their ability to predict performance from assessment reports; only three of the correlations were significant at the .05 level.

The fitness reports used in this study required the evaluator not only to give a narrative appraisal but to rate the overall performance of each of his subordinates on a 5-step adjectival scale: weak, adequate, strong, proficient, outstanding. In this study, the adjectival ratings were not made available to the judges since it was thought that differences in rating might reflect variations in

TABLE 3
PERFORMANCE AS A FUNCTION OF
ASSESSMENT PREDICTION

| Assessment prediction | Performance evaluation | |
|-------------------------------|------------------------|---------------|
| | Average or above | Below average |
| Average or above ^a | 71% | 29% |
| Below average ^b | 40% | 60% |

^a N = 17.
^b N = 15.

TABLE 4

CORRELATIONS BETWEEN INDIVIDUAL ASSESSMENT-
REPORT PREDICTIONS AND COMPOSITE
FITNESS-REPORT EVALUATIONS

| Judge | Correlations between individual ratings of assessment reports & composite (7 judges) fitness-report ratings |
|-------|--|
| 1 | .29 |
| 2 | .30* |
| 3 | .30* |
| 4 | .19 |
| 5 | .41* |
| 6 | .22 |
| 7 | .13 |

* $p < .05$, one-tailed test.

rating bias of raters more than variations in performance. Table 5 presents data indicating that the judges in this study evaluated the narrative section of the ratee's fitness reports in the same direction as the overall letter ratings assigned to each man by his supervisor. Remembering that the larger the numerical rating an individual received the higher was his judged effectiveness, individuals receiving overall "strong" ratings were judged more effective than those receiving overall "proficient" ratings ($p < .07$). The biserial correlation between the judged composite rating of effectiveness and the overall letter rating was .34. More important than the agreement of supervisors' ratings of over-

TABLE 5

MEAN EFFECTIVENESS RATINGS FOR INDIVIDUALS
RECEIVING STRONG AND PROFICIENT OVERALL
FITNESS-REPORT EVALUATIONS

| Item | Mean composite effectiveness rating ^a |
|--|---|
| Individuals receiving overall strong fitness-report evaluations ^b | 22.3 |
| Individuals receiving overall proficient fitness-report evaluations ^c | 19.1 |

Note.—An evaluation of "strong" was superior to "proficient" in the fitness-reporting system.
^a As judged by seven psychologists from fitness-report narratives only.
^b $N = 19$.
^c $N = 13$.

all performance with the judges' ratings based on the supervisor's narrative evaluation is the fact that the judges' ratings provide a greater range than is usually obtained with fitness reports in which the majority of supervisors generally restrict themselves to about two categories, as they did in this study where all the overall ratings were either proficient or strong. The high reliability of the 5-point ratings made by the psychologists suggests that a greater range of performance among personnel is recognized by supervisors than is typically reflected in their overall ratings in fitness reports.

Trait Prediction

In this portion of the study, the seven psychologists, on the basis of assessment reports only, rated all 32 employees on 25 traits or dimensions that had been abstracted from the fitness reports of the total group of individuals. Using the specially designed Trait Rating Sheet, judges predicted whether each individual would be discussed favorably or unfavorably on each dimension in his fitness report, assuming that he would be discussed on all dimensions.

A major difficulty with these data arose because 88% of the 188 statements abstracted from the fitness reports of the 16 individuals were favorable. Similarly, 74% of the total number of predictions made by the judges were positive. These high-positive base rates insured a great deal of agreement between predictions based on assessment reports and statements drawn from fitness reports. In fact, 74% of the total group of over 1,300 predictions made by the seven psychologists were "correct," that is, in agreement with the fitness-report narratives. Given the high rate of positive statements in fitness reports and the nearly as high rate of positive predictions made from assessment reports, were the psychologists able to make a significant improvement over the base rates in their prediction of these specific dimensions of performance?

One way of answering this question is presented in Table 6. If psychologists are able to predict specific dimensions of performance to a degree exceeding that which would be expected by base rates alone, then their predictions for those individuals described posi-

tively in fitness reports on a specific dimension should exceed the overall (or base rate) prediction for all persons on that dimension. Since for most dimensions the distribution of the psychologists' predictions was skewed, the median rather than the mean percentage of psychologists' predictions of favorable fitness-report descriptions on a given dimension was taken as the base rate for that dimension. For example, if 85% of the judges predicted that a certain individual would be described favorably on a given dimension and in fact he was described favorably in his fitness report on this dimension, this would constitute a successful prediction if the median percentages of judges rating all individuals positively on that dimension was 71. If, however, only 57% of the judges predicted that this person would receive favorable mention on this dimension, this would be classified as an unsuccessful prediction since it is below the 71% base rate. But if this person's fitness report had made an *unfavorable* comment about his initiative and resourcefulness, the first prediction (where 85% of the judges predicted a favorable description) would have been classified as unsuccessful since it was above the base rate while the second prediction would be successful (since only 57% of the judges predicted a favorable description of this dimension as compared with a base rate of 71%). This is a rather rigorous test, for it assumes that people mentioned favorably in their fitness reports on a specific dimension are actually stronger, and the people mentioned unfavorably, weaker on that dimension than people not mentioned one way or the other. The typical fitness report, of course, does not provide a comprehensive or systematic picture of a person's strengths or weaknesses.

Table 6 shows that for 83 of the total group of 150 positive statements drawn from fitness reports, the group of seven psychologists made predictions on the corresponding dimensions that were more in the correct (or favorable) direction than the average of the total group of predictions made on these dimensions. Similarly, for the 21 negative statements drawn from the fitness reports, the psychologists made 16 correct predictions on the corresponding dimensions. Thus, for a

TABLE 6

NUMBER OF SUCCESSFUL AND UNSUCCESSFUL PREDICTIONS MADE ON SPECIFIC PERFORMANCE AND PERSONALITY DIMENSIONS DESCRIBED IN FITNESS REPORTS

| Dimension | Successful predictions | Unsuccessful predictions | Total |
|-----------|------------------------|--------------------------|-------|
| Positive | 83 | 67 | 150 |
| Negative | 16 | 6 | 21 |
| Total | 99* | 72 | 171 |

Note.—"Successful" and "unsuccessful" were defined in terms of base rates; a successful prediction for an individual on a given dimension was recorded when the percentage of judges rating that individual in the same direction as the fitness report's narrative exceeded the median percentage of the judges rating all individuals on that dimension. (See the text for a complete description of this method.)

* $p < .02$ that this split is significantly different from a .50 : .50 split.

combined total of 99 of 171 predictions, the psychologists achieved more accurate predictions than would have been expected through base rates alone. A binomial test indicates that this ratio of successful to unsuccessful predictions exceeds a .50 : .50 (chance) split at the .02 level. (Seventeen positive statements drawn from fitness reports could not be classified as successful or unsuccessful predictions since the percentage of psychologists predicting a favorable fitness-report description fell at the median for all Ss on those dimensions.)

Because of the relatively few individuals discussed on each of the various dimensions of the Trait Rating Scale in the fitness reports (no more than 20 of 32 individuals were cited on any single dimension), it is not possible to compare the relative predictive effectiveness of the group of psychologists on different dimensions. However, there is evidence that the psychologists in this study were better able to predict weaknesses than strengths. On positive dimensions, 55% of the psychologists' predictions were successful (i.e., better than the base rates). On negative dimensions, 76% of their predictions were successful. The difference between these proportions was significant at the .05 level.

DISCUSSION

On the basis of this study, it is reasonable to conclude that psychologists can predict

significantly better than chance both overall competence and specific performance and personality characteristics of employees using only completed assessment reports prepared 1-4 yr. earlier.

The modest relationships that emerged for the prediction of overall as well as specific dimensions of effectiveness are probably artificially low, since the least promising individuals were not employed at all. This type of restriction of range is unavoidable in studies of this nature. Had it been possible to gather feedback data on all individuals assessed, it is likely that the predictive effectiveness of the psychologists would have been enhanced.

It was found that the pooled judgments of several judges yielded greater predictive accuracy than the judgments of individual psychologists. Only one of the seven judges was able to exceed the predictive accuracy of the composite judgments. As Kelley and Thibaut (1954) point out, pooling independent judgments should always enhance validity except in the situation where the judgments of the average individual correlate zero with the criterion.

The finding that psychologists were able to predict specific performance dimensions and personality characteristics better than the base rate was encouraging. It should be remembered that these predictions were made on the basis of secondary information; that is, the psychologists who made the predictions used assessment reports that were not formulated specifically toward making predictions on these dimensions. Therefore, the psychologists in this study were forced to "read between the lines" to make predictions on most of the dimensions for most of the employees. Higher predictive accuracy could be expected if the psychologists who made the predictions conducted the initial assessments with these dimensions in mind.

The finding that psychologists were better able to predict weaknesses than strengths is provocative. If substantiated by further research, it has interesting implications for the assessment process.

That psychologists can reliably generate 5-point evaluations of fitness reports that originally fell in only two categories is note-

worthy. One of the difficulties in using many standard fitness reports or appraisal ratings as criteria of job performance is their limited variance. The results of this study indicate that job-performance variance can be meaningfully expanded through a modified *Q* sort that forces reviewers of these reports to make more discriminations among individuals.

Finally, studies similar to the present one should be conducted with persons other than psychologists making predictions on the basis of assessment reports. This would be more nearly analogous to the situation at present where the psychologist, through his assessment report, supplies a consultative function to another individual (or group of individuals) who combines this report with other information in order to arrive at a selection decision. Implicit in this decision is the prediction of how well a given individual will "work out," or even whether he will "work out" at all. In the last analysis, these predictions made by the persons who typically select or reject are the most meaningful ones, and hence should be the focus of systematic study.

Meanwhile, this study does provide reassurance that the assessment process can result in meaningful predictions of job behavior as evaluated from fitness reports.

REFERENCES

- ALBRECHT, P. A., GLASER, E. M., & MARKS, J. Validation of a multiple assessment procedure for managerial personnel. *Journal of Applied Psychology*, 1964, **48**, 351-360.
- BRAY, D. W., & GRANT, D. L. The assessment center in the measurement of potential for business management. *Psychological Monographs: General and Applied*, 1966, **80**(17, Whole No. 625).
- CAMPBELL, J. T., OTIS, J. L., LISKE, R. E., & PRIEN, E. P. Assessment of higher-level personnel: II. Validity of the overall assessment process. *Personnel Psychology*, 1962, **15**, 63-74.
- DICKEN, C. F., & BLACK, J. D. Predictive validity of psychometric evaluations of supervisors. *Journal of Applied Psychology*, 1965, **49**, 34-37.
- HILTON, A. C., BOLIN, S. F., PARKER, J. W., JR., TAYLOR, E. K., & WALKER, W. B. The validity of personnel assessments by professional psychologists. *Journal of Applied Psychology*, 1955, **39**, 287-293.
- HOLTZMAN, W. H., & SELLS, S. B. Prediction of flying success by critical analysis of test protocols. *Journal of Abnormal and Social Psychology*, 1954, **49**, 485-490.
- KELLEY, H. H., & THIBAUT, J. W. Experimental studies of group problem solving and process. In G. Lindzey (Ed.), *Handbook of social psychology*.

Vol. 2. *Special fields and applications*. Cambridge: Addison-Wesley, 1954.

KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: University of Michigan Press, 1951.

OSS Assessment Staff, *Assessment of Men*. New York: Rinehart, 1948.

TAFT, R. Multiple methods of personality assessment. *Psychological Bulletin*, 1959, **56**, 333-352.

TRANKELL, A. The psychologist as an instrument of prediction. *Journal of Applied Psychology*, 1959, **43**, 170-175.

(Received December 30, 1968)

We are pleased to announce that Edwin A. Fleishman will become editor of the *Journal of Applied Psychology*, effective January 1, 1971. He succeeds Kenneth E. Clark who has been editor of the journal since 1961. Due to publication lag, manuscripts submitted after January 15, 1970 should be sent to the Editor-Elect:

Edwin A. Fleishman, Director
American Institutes for Research
8555 Sixteenth Street
Silver Spring, Maryland 20910

PSYCHODYNAMICS OF INTRINSIC AND EXTRINSIC JOB ORIENTATION

S. D. SALEH¹

AND

T. G. GRYGIER

University of Waterloo

University of Ottawa

Personality correlates of intrinsic job orientation (IO) and extrinsic job orientation (EO) are studied using a sample of 136 employees in an organization that provides social services. The results are presented for general intrinsic orientation as well as for specific factors included in the two broad categories. The results tend to indicate that concern with intrinsic factors signifies approach tendencies, while concern with extrinsic factors points to avoidance tendencies.

Classifying job factors into intrinsic and extrinsic categories has been emphasized ever since Herzberg, Mausner, and Snyderman (1959) published their book *The Motivation to Work*. Intrinsic factors are defined as those directly related to the actual performance of the job (i.e., achievement, responsibility, nature of work, etc.), while extrinsic factors are defined as those related to the environment in which the job is being performed (i.e., company policy, working conditions, interpersonal relationships, security, etc.).

Most of the recent work in this area concentrated on supporting or refuting the proposition that these two categories are two dimensions of job attitude (Burke, 1966a; Ewen, Smith, Hulin, & Locke, 1966; Graen, 1966; Herzberg, 1965; Myers, 1964; Schwartz, Jenusaitis, & Stark, 1963). A few studies, however, related the two categories to job level (Gurin, Vernoff, & Feld, 1960; Porter, 1962, 1964), age (Saleh, 1964), and sex (Burke, 1966b; Centers & Bugental, 1966; Saleh & Lalljee, in press). In the area of mental health, Hamlin and Nemo (1962) in a sample of schizophrenics found that "motivation seekers" or the intrinsically oriented improved more than the "hygiene seekers" or the extrinsically oriented.

The authors are not aware of any studies that directly investigated the relationship between the intrinsic-extrinsic dichotomy and personality variables. An analysis of such relationships should provide more insight into the nature of these categories. The present study is an attempt to fill part of this gap

by examining the personality correlates of the intrinsic orientations (IO) and the extrinsic orientations (EO), and of each specific factor included in the two broad categories.

METHOD

The study was conducted in an organization whose primary function is to provide correctional and social services to children and adolescents. The sample consisted of 136 of the technical staff, all on the same job level, who did not have formal education beyond high school. The mean age of the group was 36.4 with a standard deviation of 8.4. Only 14 Ss were female. Two scales were administered: the Job Attitude Scale (JAS; Saleh, 1964) and the Likes and Interests Test (LIT; Grygier, 1956).

The JAS consists of 16 job-related statements, each paired with each of the other 15 in a forced-choice format. Six of the statements present intrinsic factors: achievement, recognition, responsibility, nature of work, advancement, growth in skill. The other 10 present extrinsic factors: working conditions, company policy, salary, security, status, technical supervision, and salary needs for family's sake, and interpersonal interactions with supervisor, subordinates, and equals. The overall job orientation was secured by scoring only those items where an intrinsic factor was paired with an extrinsic one (60 items). By giving 1 point whenever the intrinsic factor is checked, the possible score range is 0-60. The range in the present study was 5-55. The internal consistency (split half) of this scale was .94. The JAS also provided a score for each factor by using the 15 items in which the factor was paired with the other 15, and the scores of these factors range from 0 to 15. The means and standard deviations of all dimensions are presented in Table 1.

The LIT is a slightly shorter version of the Dynamic Personality Inventory (Grygier, 1960), which is based on the psychoanalytic theory and started as a modification of the Krout-Tabin Personal Preference Scale (Krout & Krout, 1951, 1954). It ran over a dozen experimental editions: Succeeding editions of the test were factorized, examined for internal consistency and repeat reliability, and

¹ Requests for reprints should be sent to S. D. Saleh, Department of Management Sciences, University of Waterloo, Waterloo, Ontario, Canada.

TABLE 1
MEANS AND STANDARD DEVIATIONS OF THE
JAS DIMENSIONS

| Dimension | <i>M</i> | <i>SD</i> |
|---------------------------------|----------|-----------|
| IO | 30.2 | 8.9 |
| Family needs (salary) | 11.8 | 3.0 |
| Achievement | 9.2 | 2.9 |
| Growth in skill | 8.9 | 3.3 |
| Security | 8.7 | 3.9 |
| Relationships with peers | 8.6 | 3.2 |
| Relationships with subordinates | 8.1 | 2.9 |
| Advancement | 7.9 | 3.3 |
| Responsibility | 7.1 | 3.3 |
| Supervision | 7.0 | 3.3 |
| Working conditions | 7.0 | 2.9 |
| Salary | 6.8 | 3.9 |
| Relationship with supervisor | 6.5 | 2.6 |
| Creative work | 6.4 | 3.7 |
| Recognition | 5.6 | 3.6 |
| Personnel policies | 4.7 | 2.6 |
| Prestige | 4.5 | 3.4 |

Note.—Abbreviations: JAS = Job Attitude Scale; IO = intrinsic job orientation.

validated and cross-validated (Grygier, 1956). It consists of 304 items and has 30 scales.

These scales are

- H Hypocrisy: self-satisfaction with own moral standards, lack of insight into own limitations, social conformity.
- Wp Passivity: liking for comfort, warmth, and mild sensual impressions.
- Ws Seclusion and introspection and their use as a defense mechanism against social anxiety.
- OA Oral aggression: pleasure in biting and crunching, liking for strong and savory foods, suggestion of free-floating aggression and anxiety about its control.
- Od Oral defense: need for guidance and reassurance, clinging attitude.
- Om Need for freedom of movement and for emotional independence.
- Ov Verbal aggression: verbally and/or intellectually aggressive and self-assertive behavior.
- Oi Impulsiveness, changeability, spontaneity, speed of reaction, emotional expressiveness, generosity, and extravagance.
- Ou Unconventionality of outlook, originality, and individuality.
- Ah Hoarding behavior, anxious possessiveness, and stubborn, clinging persistence.
- Ad Attention to details: orderliness, conscientiousness, and perfectionism.
- Ac Conservatism, rigidity, tendency to stick to routine.
- Aa Submissiveness to authority and order.
- As Anal sadism: emphasis on strong authority, cruel laws, and discipline.

- Ai Insularity: reserve and mistrust, social and racial prejudice.
- Pn Narcissism: concern with clothes and appearance; sensuous enjoyment of luxury.
- Pe Exhibitionism: conscious enjoyment of attention and admiration
- Pa Active Icarus complex; psychophysical drive, drive for achievement.
- Ph Fascination by height, space, and distance (passive Icarus complex): aspirations at the fantasy level.
- Pf Fascination by fire, wind, storms, and explosions (sensual aspects of the Icarus complex): perceptiveness of sensual impressions, vivid imagination.
- Pi Icarian exploits: interest in active, pioneering exploration and a liking for adventure.
- TI Enjoyment of tactile impressions: interest in handicrafts and creative manipulation of objects.
- CI Creative, intellectual, and artistic interests.
- M Masculine sexual identification and tendency to adopt masculine social roles, interests, and attitudes.
- F Feminine sexual identification and a tendency to adopt feminine social roles, interests, and attitudes.
- MF Tendency to seek social roles (irrespective of their masculine or feminine characteristics).
- SA Interest in social activities.
- C Interest in children, need to give affection.
- EP Ego defensive persistence: tendency to react with renewed effort in the face of difficulties or opposition.
- EI Initiative, self-reliance, and a tendency to plan, manage, and organize.

The general job orientation scale (IO) and all the subscales of the JAS were correlated with the 30 scales of the LIT.

RESULTS AND DISCUSSION

Table 2 shows that the intrinsically oriented Ss scored low on two of the personality variables (Od and Ac), which means that they were relatively more independent, confident, and flexible. They did not feel that they needed much guidance or reassurance (Od), and they rejected stability, routine, and conventional standards (Ac).

Moreover, the IO indicates a tendency to counteract with renewed effort in the face of difficulties or opposition, "doubling efforts after criticism," "sticking at a job when no results are forthcoming," "concentrating on one task for a long time" (EP). Initiative, self-reliance, and leadership (EI) are also some of their qualities. They have a strong tendency to be independent and to seek freedom of movement (OM). They would, for instance, "take risks," "start out on

TABLE 2

SIGNIFICANT CORRELATION BETWEEN

| Dimension | H | Wp | OA | Od | Om | Ov | Oi | Ou | Ac | Aa | As | Ai |
|--------------------------------|------|-----|------|------|------|------|------|------|------|-----|------|------|
| Intrinsic orientation | | | | -.20 | .22 | | | | -.26 | | | |
| Achievement | .30 | | | .18 | | | | | | | | |
| Creative work | | | | | .17 | | | | -.22 | | | |
| Advancement | | | | -.28 | | .28 | | | | | | |
| Responsibility | | | | | | .19 | | | | | | |
| Growth in skill | | | | | | | | | -.22 | | | -.17 |
| Security | | | -.21 | .30 | -.32 | -.23 | | -.26 | .40 | .19 | | .27 |
| Salary | -.37 | | | | | | .23 | | | | | |
| Family needs (salary) | -.31 | .20 | | | | | | | | | | |
| Relationship with subordinates | | | | | | -.18 | -.22 | | | | -.18 | -.23 |
| Relationship with supervisor | | | | | | -.25 | .17 | | | | | |
| Personnel policies | | | | | | | | | | | | -.23 |
| Prestige | | | | | | | | | | | | |

Note.—*N* = 136. A correlation of .17 is significant at the .05 level, and a correlation of .23 is significant at the .01 level. Abbreviations: JAS = Job Attitude Scale; LIT = Likes and Interests Test. See text for scale abbreviations.

new ventures,” and “have no permanent attachments.”

As would be expected in a group whose primary function is to provide social service, the IO in our sample expressed a tendency to seek social roles (MF), especially masculine roles (M), and to be more interested in social activities (SA) than the EO.

The results also show that the IO tend consciously to enjoy attention and admiration and desire to seek prominence (Pe). For instance, they would enjoy “being a Master of Ceremonies,” sitting in the front row at a meeting,” and “appearing on the stage.”

The last significant correlation (PL) indicates fascination by height, space, and distance and is related to the flow of ideas and imagination of IO Ss.

The correlations in Table 2 show that “creative work” is the dimension most similar to the IO dimension. The Ss who were concerned with the nature of their job shared the following characteristics with the intrinsically oriented. Both dislike rigidity of approach (AC), have a need for emotional independence (OM), are imaginative (PH), have a strong *n*Achievement (Pa), and both seek social roles (MF), especially masculine ones (M). The M scale has such items as “drawing up plans introducing new ideas” and “making new gadgets and mechanical devices.”

The results suggest that the psychodynamics of advancement and of responsibility are similar. Both are characterized by self-assertion (Ov), drive for achievement (Pa), and a tendency to seek masculine social roles (M). However, concern with advancement correlates negatively with the tendencies to seek to give affection (Od, C), while responsibility is characterized by assuming leadership roles (EI). It is of interest to note that three of the four correlations that describe responsibility (Pa, M, EI) describe the prestige dimension. The first two correlations are also shared with advancement. This similarity suggests considering prestige an intrinsic factor rather than an extrinsic one.

In contrast with the need for advancement and responsibility, emphasis on relationships with subordinates is negatively related to expression of one’s own individuality and self-assertion (Ou), to the drive for achievement (Pa), and to initiative and self-reliance (EI). Also the negative correlation with the Pn scale suggests a denial of self-concern and of narcissistic needs.

The salary dimension as well as the dimension of family need for money seem to be, in general, negatively related to different kinds of activity (H, MF, SA, EI, TI), while they are positively related to passive enjoyment

THE JAS AND THE LIT

| Pn | Pe | Pa | Ph | Pf | Pi | TI | CI | M | F | MF | SA | C | EP | EI |
|------|------|------|------|------|------|------|------|------|-----|------|------|------|------|------|
| | .24 | .31 | .19 | | | .20 | | .27 | | .21 | .19 | | .17 | .28 |
| | | .23 | | | | | .20 | .21 | .21 | .27 | | | .24 | |
| | | .23 | | | | | | .19 | | .23 | | | | |
| | | .27 | | | | | | .25 | | | | -.18 | | .31 |
| | -.23 | -.28 | -.29 | -.24 | -.17 | | -.21 | -.22 | | | -.22 | .19 | -.17 | -.26 |
| | | | | | | -.24 | -.18 | | | -.25 | -.26 | -.20 | -.22 | -.24 |
| | | | | | | | | | | -.18 | -.20 | | | -.18 |
| -.20 | | -.22 | | | | | | | | | | | | -.20 |
| | | .18 | | .18 | .18 | | | .19 | | | | | | .18 |

of comfort, generosity, and extravagance (Wp, Oi).

The security dimension appears to be quite meaningful psychologically since it is correlated with the largest number of the LIT scales. It is of interest to note that all of the scales that describe the IO, with only one exception (FM), are correlated with an opposite sign with security. Those positively correlated with IO are negatively correlated with security and vice versa.

One general observation about the results is that the majority of correlations with the intrinsic factors were positive, while with the extrinsic factors the majority were negative. It should be emphasized that the difference is not only mathematical but also is psychologically meaningful. This seems to indicate that concern with intrinsic factors signifies approach tendencies while the regard for extrinsic factors is characterized by avoidance tendencies.

In this regard the results support Herzberg's notion of considering the "motivators" different in nature from the "hygienes," and they also show the meaningfulness of differentiating between the two dimensions.

In conclusion it should be pointed out that although the present study has presented some meaningful results, more investigations, on

different population samples, are needed to make any adequate generalization.

REFERENCES

BURKE, R. J. Are Herzberg's motivations and hygienes unidimensional? *Journal of Applied Psychology*, 1966, 50, 317-321. (a)
BURKE, R. J. Differences in perception of desired job characteristics of the same sex and the opposite sex. *The Journal of Genetic Psychology*, 1966, 169, 37-46. (b)
CENTERS, R., & BUGENTAL, D. E. Intrinsic and extrinsic job motivations among different segments of the working population. *Journal of Applied Psychology*, 1966, 50, 193-197.
EWEN, R. B., SMITH, P. C., HULIN, C. L., & LOCKE, E. A. An empirical test of the two-factor theory. *Journal of Applied Psychology*, 1966, 50, 544-550.
GRAEN, G. Motivator and hygiene discussions for research and development engineer. *Journal of Applied Psychology*, 1966, 50, 563-566.
GRYGIER, T. G. *The Likes and Interest Test*. London: National Foundation for Educational Research in England and Wales, 1956.
GRYGIER, T. G. *The Dynamic Personality Inventory*. London: National Foundation for Educational Research in England and Wales, 1960.
GURIN, G., VERNOFF, J., & FELD, S. *Americans view their mental health*. New York: Basic Books, 1960.
HAMLIN, R., & NEMO, R. Self-actualization in choice scores of improved schizophrenics. *Journal of Clinical Psychology*, 1962, 18, 51-54.
HERZBERG, F. The motivation to work among Finnish supervisors. *Personnel Psychology*, 1965, 18, 398-402.
HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. *The motivation to work*. New York: Wiley, 1959.

- KROUT, M. H., & KROUT, J. *Personal Preference Scale*. Chicago: Chicago Psychological Institute, 1951.
- KROUT, M. H., & KROUT, J. Measuring personality in developmental terms: The personal preference scale. *Genetic Psychology Monograph*, 1954, **50**, 289-335.
- MYERS, S. M. Who are your motivated workers? *Harvard Business Review*, 1964, **42**, 73-78.
- PORTER, L. W. Job attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology*, 1962, **46**, 375-384.
- PORTER, L. W. *Organizational patterns of managerial job attitudes*. New York: American Foundation for Management Research, 1964.
- SALEH, S. D. A study of attitude change in the pre-retirement period. *Journal of Applied Psychology*, 1964, **48**, 310-312.
- SALEH, S. D., & LALLJEE, M. Sex and job orientation. *Personnel Psychology*, 1969, in press.
- SCHWARTZ, M. M., JENUSAITIS, E., & STARK, H. Motivational factors among supervisors in the utility industry. *Personnel Psychology*, 1963, **16**, 45-53.

(Received October 24, 1968)

COMPARISON OF SEVERAL PATTERNS OF COMMUNICATION

CHARLES J. O'LEARY AND MORTON GOLDMAN¹

University of Missouri, Kansas City

Four patterns of communication were compared with respect to the amount of material recalled. Pattern 1 consisted of a "receiver" listening to one "transmitter." Pattern 2 was similar, except that the "receiver" listened to two "transmitters" speaking simultaneously. Pattern 3 required the "receiver" to talk while listening to one "transmitter." Pattern 4 had the receiver talk at the same time that he listened to two simultaneous "transmitters." Statistically significant results were obtained favoring the "receiver" who listened to one rather than two transmitters (even though in the case of two "transmitters" twice the number of different facts were heard) and listened rather than talked and listened. A significant interaction also occurred. Information retained by the "receiver" from each of the two simultaneous "transmitters" was compared.

In studying the interrelation between individuals in any social system the general area of communication is an important concern. Dealing with communication becomes very critical in the study of small collectivities or groupings in which success of goal attainment is directly proportional to efficiency of information transmission, receipt, and retention.

Casual observation very readily points out that in what is referred to as "everyday interaction" there exists a very definite pattern of "communicative acts." This patterning will be referred to as the "normal pattern." The normal pattern is one in which the communicative acts are sequential in nature with little or no superimposition. If one person is speaking and another starts and persists, the first individual stops. Thus there is a dichotomous positioning of speaker and listener.

In terms of efficiency, this patterning would appear to be less than optimal. Yet while Miller (1965, p. 95) points out that "there is no a priori reason why two people . . . could not question and answer simultaneously," a search of the literature of different patterns of communication did not yield any research. However, some related studies have been conducted where Ss were presented with overlapping messages to more than one sense organ. Also related are studies that deal with selective perception examining the conditions under which more attention is paid to one

message than another, where both messages are heard simultaneously. Studies in these areas have been reviewed by Horn (1965) and by Broadbent (1958).

Many variations of verbal communication other than the normal pattern are possible. An individual can listen to several people speaking simultaneously; and an individual can speak at the same time that he is listening to others. In an initial study, it was decided to study these two factors using a 2×2 design. The receiving Ss ("receivers") hear statements from either one individual or two individuals speaking simultaneously and each presenting different facts ("transmitters"). Under each of these two conditions the receiver will listen only, or will be talking at the same time that he is listening to the transmitters impart their facts. Hence one factor of the current study deals with listening to different numbers of transmitters (one or two), and the other factor deals with the behavior of the receiver (listening only, or talking and listening). In all treatments the dependent variable will be the number of facts the receiver can recall.

Related to the talking and listening procedures to be examined in the current study is a technique called the shadowing method, which has been employed in studying selective listening. The shadowing method, first described by Cherry (1953), has an S repeat aloud a spoken prose message as it is heard with as short an interval as possible between the spoken and shadow message. However, in

¹ Requests for reprints should be sent to Morton Goldman, Psychology Department, University of Missouri, Kansas City, Missouri 64110.

the current paper, Ss were asked to say aloud a message differing from the one to which they were listening.

Some of the results of the current study might appear obvious. For example, it might be anticipated that the amount of information that the receiver can recall would be greater when he devotes all of his attention to listening than if he diverts his attention by talking and listening at the same time. However, in the treatment where the receiver is listening simultaneously to two transmitters, each presenting different material, the number of facts presented in a given time period is twice the number presented to the receiver who is in the treatment where there is only one transmitter. How would the amount of information recalled by the receivers compare for these two treatments? The results by no means are obvious. Further the design to be used allows for an answer to the question of when the receiver hears two transmitters, whether he pays equal attention to both, or disregards one transmitter in favor of the other.

Since the communication patterns to be used in this study generally can be assumed to be unusual for most individuals, it was decided to give Ss five trials to see if with repeated practice results would change.

METHOD

Subjects. The Ss used in the study were recruited from the General Psychology course at the University of Missouri, Kansas City. Ten Ss were randomly assigned to each of the four treatments making a total of 40 Ss.

Treatments. Four treatments described above were devised: (a) the NT-L₁ treatment consisted of having an S (not talk) act as a receiver and listen to

one person, a transmitter, impart information; (b) the NT-L₂ treatment was similar to the above, except that the receiver listened to two transmitters speak simultaneously, each imparting different information; (c) the T-L₁ treatment required the receiver to talk at the same time he was listening to one transmitter—the message spoken by the receiver was different than the message spoken by the transmitter; (d) the T-L₂ treatment required the receiver to talk while he listened to two transmitters simultaneously talking, each message being different. Thus, in the T-L₁ treatment two people are speaking at the same moment, one receiver and one transmitter; in the T-L₂ treatment three people are speaking at the very same moment, one receiver and two transmitters. In all the treatments, the receiver was given five trials, each trial consisting of new transmitters imparting different information. The transmitters consisted of college students who were not acquainted with Ss serving as receivers.

Procedure. Each transmitter was asked to complete a form requesting the following 10 items: (a) name, (b) favorite sport, (c) college major, (d) advisor's name, (e) mother's first name, (f) father's first name, (g) father's occupation, (h) possible future occupation, (i) religious affiliation, and (j) place of birth. After being placed in a different randomized order for each S these 10 items served as the information which each transmitter was to recite to the receivers. To facilitate reciting the items, the transmitters were given a short period of time to familiarize themselves with the order of the items. The information forms were kept in front of them during the transmission. This procedure was also followed for the receivers in the T-L₁ and T-L₂ treatments, where they had to recite information. The Ss all spoke at approximately the same speed (the speed of normal conversation) and generally took a similar amount of time to impart the items of information. In all four treatments the receivers, upon the completion of each trial, immediately recorded the information they were able to retain. This retained information, recorded by the receivers, served as the dependent variable.

RESULTS

In each of the four treatments the transmitters imparted 10 factual statements about themselves. At the end of each trial the receivers were asked to write down as many facts as they could recall. One point was allotted to a receiver for each correct recorded fact. Table 1 shows the means for the 10 receivers in each of the four treatments over the five trials.

The results were analyzed with a 2 × 2 trend analysis of variance. For the factor comparing the amount of information retained by

TABLE 1
MEAN RETENTION FOR TRIALS OF EACH TREATMENT, TRIALS OVER ALL TREATMENTS, AND TREATMENTS OVER ALL TRIALS

| Treatments | Trials | | | | | Treatments average for trials |
|----------------------------|--------|------|------|------|------|-------------------------------|
| | 1 | 2 | 3 | 4 | 5 | |
| NT-L ₁ | 8.1 | 8.8 | 8.3 | 6.8 | 7.5 | 7.59 |
| NT-L ₂ | 6.8 | 6.0 | 8.0 | 6.8 | 8.2 | 7.16 |
| T-L ₁ | 5.1 | 5.0 | 4.5 | 5.0 | 4.2 | 4.84 |
| T-L ₂ | 3.4 | 3.7 | 3.3 | 2.7 | 2.6 | 3.14 |
| Trials over all treatments | 5.85 | 5.88 | 6.02 | 5.32 | 5.62 | |

the receiver when he listened to one transmitter as opposed to two transmitters giving simultaneous information, an $F = 8.02$ ($df = 1/36$, $p < .01$) was obtained, favoring listening to one transmitter. This result occurred even though the receiver, when he listened to one transmitter, heard only 10 facts and thus could obtain a maximum of 10 points; while the receiver, when he listened to two simultaneous transmitters, heard 20 facts and could attain a maximum of 20 points. As would be anticipated, for the factor comparing retained information when the receiver listened only as opposed to talking while listening, an $F = 65.48$ ($df = 1/36$, $p < .001$) was obtained, favoring listening only. For the interaction of the above two factors an $F = 11.18$ ($df = 1/36$, $p < .01$) was obtained which reflected that Ss in the T-L₂ treatment recalled fewer facts in comparison to the T-L₁ treatment than did Ss in the NT-L₂ treatment in comparison to the NT-L₁ treatment. Stated another way, a greater drop in recall occurred for Ss listening to one transmitter when Ss themselves were also talking rather than only listening.

No discernible pattern could be detected for any of the treatments over the five trials, the F for the trial source of variance being less than 1. The F s obtained for the interactions of the trials and the other factors were non-significant.

In the two treatments, NT-L₂ and T-L₂, where Ss are listening to two transmitters simultaneously, the question can be raised as to whether the receivers recall approximately the same amount of material from each transmitter or if Ss tend to favor one of the two transmitters and slight the other. Further, an additional question can be asked—if this favoring process occurs differentially in the T-L₂ treatment as compared to the NT-L₂ treatment. To obtain evidence bearing on these two questions, for each receiver (combining all trials in the NT-L₂ and T-L₂ treatments), the transmitters were divided into two groups in the following manner. Of the two transmitters speaking simultaneously to a given receiver, the transmitter from whom that given receiver recalled more information was called the high-recall transmitter, and

TABLE 2
MEAN RECEIVERS RECALL SCORES FOR HIGH AND LOW
TRANSMITTERS IN THE NT-L₂ AND T-L₂
TREATMENTS COMBINING THE
FIVE TRIALS

| Receivers | Score |
|-------------------|-------|
| NT-L ₂ | |
| High recall | 22.9 |
| Low recall | 12.3 |
| T-L ₂ | |
| High recall | 12.1 |
| Low recall | 3.4 |

the transmitter from whom the receiver recalled less information was called the low-recall transmitter. The amount of information recalled by the receivers from the high- and low-recall transmitters are presented in Table 2.

In the NT-L₂ treatment the receivers obtain approximately 2/3 of their score from the high-recall transmitter and 1/3 from the low-recall transmitter. In the T-L₂ treatment, the receivers obtain approximately 4/5 of their score from the high-recall transmitter and 1/5 from the low-recall transmitter. Comparing the high-recall mean with the low-recall mean in each of the two treatments, significant t 's were obtained for both treatments ($p < .001$). Comparing the mean percentage difference between the high- and low-recall score in the NT-L₂ treatment with mean percentage difference between the high- and low-recall score in the T-L₂ treatment, a significant t was also obtained ($p < .01$). Thus, when a receiver is listening to two transmitters, he recalls significantly more information from one than from the other. He does not recall approximately equal amounts of information from each transmitter. Further, the percentage recall difference between the two transmitters is further increased if the receiver is also talking rather than only listening.

DISCUSSION

In the current study it was found that the amount of information recalled when listening to one transmitter is significantly greater than when listening to two transmitters speaking simultaneously. This result occurred in

spite of the fact that when two transmitters spoke simultaneously, 20 facts were being presented, as compared to 10 facts when only one person spoke. When the receivers were listening to one transmitter, they were able to recall approximately 75% of the material, and thus even in this condition were unable to process all the information. In the situation when two transmitters were used, the receivers were listening to twice the amount of material, but in the same unit of time. The results showed that further overloading the receivers past the point of maximum operating ability leads to a decline in absolute efficiency.

Another finding of the present study is that speaking at the same time that one is listening significantly reduces the amount of material that can be recalled over not speaking and listening. Since in this study, the talking receivers were functioning at slightly less than 50% efficiency in recall of the material presented to them, it would be anticipated that there would be an even larger relative deterioration than for nontalking receivers, when two transmitters were used. This proved to be the case as supported by the obtained significant interaction.

In the NT-L₂ treatment, each receiver heard 20 facts in a given unit of time as compared to 10 facts for the receivers in the NT-L₁ treatment. When a new trial was presented, different transmitters were used which gave new facts. It was found that with repeated trials, at least for five repetitions, the results within chance fluctuations did not change. It would have been possible in two ways to arrange the conditions of the experiment so that the number of facts per unit of time would have remained the same for the NT-L₁ and NT-L₂ treatments: (a) the speaking speed of the transmitters in the NT-L₂ treatment could be made twice as long; or (b) the speaking speed of the transmitters could remain the same but the transmitters in the NT-L₂ treatment could have repeated the same information for two trials. The comparison of the amount of recall of the receivers in the NT-L₂ and NT-L₁ treatments under these two conditions could be explored in further research. This same comparison could

also be made for the T-L₁ and T-L₂ treatments.

The results of the study showed that when one individual is listening to two simultaneous speakers, he does not pay the same attention to each. Further, if the listener is himself talking, an even greater proportion of his attention is given to one of the two other talking individuals. Would this result have occurred to an even larger extent if NT-L₃ or T-L₃ treatments had been used? Still an additional area of further research would be to investigate the factors which determine to which of the two speakers a given listener will devote more of his attention.

Reference above has been made to the shadowing procedure used by Cherry (1953). Cherry had his Ss shadow a voice presented to one ear while a different unrelated message appeared at the other ear. The results showed that Ss could attend to the required message while ignoring the unrelated message. A modification of shadowing occurs when a translator converts a message from one language into another at the same time he is listening to the message to be translated. In this case, the simultaneous translation or shadow message is not a duplication of the same sounds and words. Treisman (1965) has been concerned with comparing shadowing and simultaneous translation. The procedure used in the T-L₂ treatment reported here can be thought of as still a further modification of shadowing. The Ss in this treatment were required simultaneously to attend to two different messages while at the same time saying aloud a third message, all three messages containing the same factual categories of information but different content, where the information spoken by S pertains to himself. Thus the current study compared the amount of recall of Ss who were exposed to several conditions of simultaneous perception. More information was retained by Ss in the NT-L₁ treatment than by Ss in the NT-L₂, T-L₁, and T-L₂ treatments. Nevertheless, Ss could recall facts from two different messages simultaneously presented while responding aloud with a related but unique third message.

REFERENCES

- BROADBENT, D. E. *Perception and communication*. New York: Pergamon Press, 1958.
- CHERRY, E. C. Some experiments on the recognition of speech, with one and two ears. *Journal of Acoustical Society of America*, 1953, **25**, 975-979.
- HORN, G. Physiological and psychological aspects of selective perception. In D. S. Lehrman, R. A. Hinde, & E. Shaw (Eds.), *Advances in the study of behavior*, Vol. 1. New York: Academic Press, 1965.
- MILLER, G. A. Speaking in general. In I. D. Steiner & M. Fishbein (Eds.), *Current studies in social psychology*. New York: Holt, Rhinehart and Winston, 1965.
- TREISMAN, A. M. The effects of redundancy and familiarity on translating and repeating back a foreign and ■ native language. *British Journal of Psychology*, 1965, **56**, 369-379.

(Received October 28, 1968)

WORK VALUES AND JOB SATISFACTION

MILTON R. BLOOD¹

University of Illinois

Two groups of airmen (students and persons in permanent assignments) completed measures of their job satisfaction and of their work values. Consistent relationships appeared between these two sets of variables. Evidence is presented which indicates that the job satisfaction variance controlled by work values is independent of that controlled by other variables.

The way a person evaluates work in general should be related to his attitudes toward his particular job. Someone who thinks that all work is an abomination to be undertaken only when all other strategies fail will likely be unhappy even in the most pleasant work situation. On the other hand, a person who feels that personal worth results only from self-sacrificing work or occupational achievement would likely derive some satisfaction even in a demanding menial position.

Previous investigators have discussed work values related to the ideals of the Protestant Ethic (Weber, 1958). Lenski (1961) reported a study which utilized a stratified sample of Detroit residents. He found differences in work values between four socioreligious groups. His most general finding related to work values was that white Protestants and Jews were more likely to be committed to the spirit of capitalism and the ideals of the Protestant Ethic than were Negro Protestants and Catholics. A similar conclusion is supported by the findings of Turner and Lawrence (1965). Among workers from rural communities who were predominantly Protestant they found job responses which would be expected from persons ascribing to the ideals of the Protestant Ethic. They found responses which would not be predicted from Protestant Ethic ideals among workers in urban areas who were predominantly Catholic.

If such differences are predictable from knowledge of religious affiliation, psychological explanation requires that they be mediated by some psychologically measurable difference. Differences in the job responses of Protestants and Catholics could be mediated

by differences in work values. If the work value differences can be measured by some psychological measurement device, it should allow the prediction of within-group differences in job responses as well as between-group differences. This study is an attempt at the measurement of individual differences in work values. It was predicted that persons who ascribe to Protestant Ethic ideals would be more satisfied with their job.

METHOD

As a part of a larger study of 448 airmen and noncommissioned officers from the United States Air Force (Blood, 1968), Ss were asked to complete the Job Description Index (JDI) scales (Smith, Kendall, & Hulin, 1969), two Faces scales (Kunin, 1955) which measured satisfaction with the job in general (JIG) and satisfaction with life in general (LIG), and an eight-item scale intended to measure amount of agreement with the Protestant Ethic. There were 420 usable questionnaires. Of these, 114 were from airmen who were enrolled as full-time students in courses in aircraft maintenance. The other 306 Ss were serving in permanent assignments on a variety of low skill level tasks principally as technicians or as maintenance, transportation, or supply workers.

The Protestant Ethic scale had four items which were intended to be in agreement with the Protestant Ethic and four which did not agree with the ideals of the Protestant Ethic. For each item, Ss responded with a number from 1 to 6 where 1 = disagree completely and 6 = agree completely. A component analysis of the eight items with a Varimax rotation of two components demonstrated that the two subsets of items were appropriately interrelated. Table 1 shows the items and their component loadings.

The four items loading heavily on the first component (Items 2, 4, 6, and 7) were summed for each individual and called the "proProtestant Ethic" score. The four items with large loadings on the second component (Items 1, 3, 5, and 8) were summed and called the "nonProtestant Ethic" score. These two work value dimensions were correlated .113 among the 114 students in the sample and

¹ Requests for reprints should be sent to the author, who is now at the Department of Psychology, University of California, Berkeley, California 94720.

TABLE 1
LOADINGS OF PROTESTANT ETHIC ITEMS ON
VARIMAX ROTATED COMPONENTS

| Items | Components | |
|--|------------|------|
| | I | II |
| 1. When the workday is finished, a person should forget his job and enjoy himself. | .13 | .53 |
| 2. Hard work makes a man a better person. | .60 | .11 |
| 3. The principal purpose of a man's job is to provide him with the means for enjoying his free time. | .12 | .61 |
| 4. Wasting time is as bad as wasting money. | .67 | .02 |
| 5. Whenever possible a person should relax and accept life as it is, rather than always striving for unreachable goals. | -.10 | .59 |
| 6. A good indication of a man's worth is how well he does his job. | .57 | .02 |
| 7. If all other things are equal, it is better to have a job with a lot of responsibility than one with little responsibility. | .58 | -.07 |
| 8. People who "do things the easy way" are the smart ones. | -.06 | .57 |

-.028 among the 306 permanent party members in the sample. Correlations were computed between the Protestant Ethic dimensions and the satisfaction measures. The five JDI scales were included separately and also summed in these analyses. Because there was evidence that there were differences between permanent party and technical school students in responses to JDI scales (Blood, 1968) the correlations were made separately for permanent party and students.

RESULTS AND DISCUSSION

Table 2 shows the correlations between the satisfaction measures and the Protestant Ethic dimensions in both the student sample and the permanent party sample. Though none of the correlations is large, the directions of the relationships are obvious. With only two exceptions the data show that agreement with the Protestant Ethic is directly related to satisfaction, and agreement with nonProtestant Ethic items is inversely related to satisfaction. This result implies that the more a worker agrees with the ideals of the Protestant Ethic, the more he will be satisfied in his work and with life in general. There

are other important influences on the satisfaction of workers. In the sample for this study there were significant differences between students and permanent party on the satisfaction variables. Nonetheless, within the student and permanent party samples, among workers on similar tasks in the same organization who share the same reward structure, additional variance in job satisfaction was predictable.

In order to assess the contribution of the Protestant Ethic dimensions to job satisfaction relative to other variables, a multiple correlation was computed for each of the satisfaction measures using age, education, tenure, father's occupation, and the Protestant Ethic dimensions as independent variables. The results of these analyses are shown in Table 3. Darlington (1968) has pointed out the dangers of overinterpreting multiple regression coefficients. However, it is possible to see that the Protestant Ethic dimensions do make a contribution to the prediction of job satisfaction, especially if we consider only the general measures of job satisfaction, JDI sum, and JIG. In addition to the regression coefficients Darlington (1968) suggests consideration of validity and usefulness² in assessing the "importance" of predictor variables. The validity coefficients for the Protestant Ethic dimensions rank 1 and 2 among all predictors

² The usefulness of a particular predictor is the difference between R^2 computed with all of the predictors and R^2 computed with all of the predictors included except the predictor of interest.

TABLE 2
CORRELATIONS BETWEEN SATISFACTION MEASURES
AND PROTESTANT ETHIC DIMENSIONS

| Satisfaction variables | Students (<i>N</i> = 114) | | Permanent (<i>N</i> = 306) | |
|------------------------|-------------------------------|--------|--------------------------------|-------|
| | Pro | Non | Pro | Non |
| JDI sum | .18* | -.24** | .16** | -.12* |
| JDI work | .09 | -.16* | .17** | -.12* |
| JDI supervisor | .06 | -.15 | .09* | -.01 |
| JDI people | .18* | -.15 | .10* | -.13* |
| JDI pay | -.02 | -.31** | .14* | -.05 |
| JDI promotion | .28** | .02 | .05 | -.06 |
| Job in general | .22* | -.17* | .10* | -.13* |
| Life in general | .08 | -.09 | .17** | -.06 |

* $p < .05$.
** $p < .01$.

TABLE 3
STANDARDIZED REGRESSION WEIGHTS AND MULTIPLE CORRELATION COEFFICIENTS FOR THE
PREDICTION OF SATISFACTION MEASURES AMONG STUDENTS AND PERMANENT PARTY

| Students (N = 114) | | | | | | | | |
|-----------------------|---------------------|------|-------|--------|------|------|------|------|
| Independent variables | Dependent variables | | | | | | | |
| | Sum | Work | Super | People | Pay | Prom | JIG | LIG |
| Age | -.06 | .04 | -.12 | .01 | -.02 | -.12 | .06 | .01 |
| Education | -.10 | -.14 | -.04 | -.12 | .01 | .04 | -.06 | -.15 |
| Tenure | .11 | .09 | .14 | -.02 | .03 | .08 | .01 | -.01 |
| Father's occ. | .05 | .03 | .12 | -.02 | -.12 | .09 | -.03 | .04 |
| Pro (PE) | .22 | .12 | .08 | .20 | .01 | .28 | .24 | .09 |
| Non (PE) | -.24 | -.16 | -.14 | -.17 | -.33 | .00 | -.20 | -.09 |
| R | .35* | .26 | .27 | .28 | .33* | .32* | .31 | .19 |

| Permanent party (N = 306) | | | | | | | | |
|---------------------------|-------|-------|------|-------|------|-------|------|------|
| Age | .16 | .05 | .20 | .14 | -.01 | .09 | -.01 | -.12 |
| Education | -.08 | -.05 | -.03 | -.10 | -.04 | -.03 | -.05 | -.01 |
| Tenure | -.06 | .08 | -.13 | .05 | .09 | -.34 | .12 | .11 |
| Father's occ. | .12 | .14 | .04 | .09 | .10 | .01 | .07 | .03 |
| Pro (PE) | .14 | .15 | .08 | .07 | .13 | .07 | .09 | .18 |
| Non (PE) | -.13 | -.13 | -.02 | -.16 | -.06 | -.05 | -.15 | -.06 |
| R | .26** | .28** | .15 | .28** | .20* | .27** | .22* | .19* |

* $p < .05$.
** $p < .01$.

for predicting JDI sum and JIG in the student sample, and they rank in the top three predictors for predicting JDI sum and JIG in the permanent party sample. In both samples the usefulness of the Protestant Ethic dimensions ranks 1 and 2 among predictors of JDI sum and JIG.

Whether a causal relationship exists between the work value dimensions and job satisfaction is a researchable question. It seems more logical to the author to assume that work values precede and influence job satisfaction rather than the opposite. Future research should investigate this relationship. Perhaps higher job satisfaction is partly a consequence of congruence between individual and institutional goals. If future research establishes that this is the case, it would constitute a justification for the suggestion that workers should hold goals similar to those of management. This is now implied by human relations theorists without justification.

Expansion and refinement of the Protestant Ethic dimension measurements should also be undertaken in future studies. Not only additional items, but other item formats should be attempted. An antiProtestant Ethic dimension should be added, and hopefully the nonProtestant Ethnic dimension should be defined in terms of what it does specify rather than what it does not specify.

As a final thought, some recent attempts to assimilate hard-core unemployed into the industrial work force have attempted to instill ideals (resocialize) similar to the ideals of the Protestant Ethic, for example, hard work brings rewards, occupational achievements bring prestige, and so forth. In evaluating the impact of these programs, it will be helpful if it is possible to measure changes in such work values. One of the first concerns of administrators of such programs should be to find out if changes in work values are accompanied by changes in job satisfaction and job performance.

REFERENCES

- BLOOD, M. R. Relationships among workers' backgrounds, attitudes, behaviors, and value systems. Unpublished doctoral dissertation, University of Illinois, 1968.
- DARLINGTON, R. B. Multiple regression in psychological research and practice. *Psychological Bulletin*, 1968, 69, 161-182.
- KUNIN, T. The construction of a new type of attitude measure. *Personnel Psychology*, 1955, 8, 65-77.
- LENSKI, G. *The religious factor*. New York: Doubleday, 1961.
- SMITH, P. C., KENDALL, L. M., & HULIN, C. L. *Measurement of satisfaction in work and retirement*. New York: Rand McNally, 1969.
- TURNER, A. N., & LAWRENCE, P. R. *Industrial jobs and the worker: An investigation of response to task attributes*. Boston: Harvard University, 1965.
- WEBER, M. *The protestant ethic and the spirit of capitalism*. (Trans. by T. Parsons). New York: Scribner, 1958.

(Received November 19, 1968)

USE OF LEADERSHIP POWERS IN INDUSTRY¹

DAVID KIPNIS² AND JOSEPH COSENTINO

Temple University

This study investigated (a) the range of corrective powers available to military and industrial supervisors when correcting subordinates' behavior and (b) the factors influencing the supervisors' use of these powers. Both situational and personal factors (number of employees supervised, years of experience as a supervisor, and the nature of the problem presented by the subordinate) were found to influence the supervisor's choice of corrective power. Military supervisors relied more on direct attempts to change subordinates' behavior through reliance upon extra instruction, direct punishments, and changes in the task environment of subordinates. Industrial supervisors relied more on their persuasive powers.

Studies of leadership behavior have focused mainly upon the identification of the dimensions of a leader's behavior and the consequence of variations in these dimensions for a subordinate's morale and productivity (Bales, 1953; Fiedler, 1965; Fleishman, 1953; Kahn & Katz, 1960). These studies have revealed the importance of leadership behavior as related to task direction and to maintaining the socioemotional well-being of subordinates. Less attention has been paid to the question of how leaders use the formal social powers associated with their organizational roles. Yet this question is of particular interest in industrial and military organizations. By virtue of their roles, formally appointed leaders control resources that are valued or required by subordinates. Among these resources are the control of sanctions, control of communication channels, and control of the direction of task performance. In essence these controls, or social powers, provide the means by which the formally appointed leader can exercise influence and thus have a central role in mediating the outcomes for subordinates.

Many questions can be raised concerning these powers. For example, does the amount of experience the individual has had as a supervisor relate to his use of the powers the organization allows him to control? In an

unpublished doctoral dissertation by Schreiber (cited in Carter, 1952) it was found that when inexperienced leaders were given too much power, their behavior was disrupted. What situational factors affect the use of social powers? Do supervisors directing large numbers of men use their powers in the same fashion as those directing fewer men? Does the personality of the supervisor influence the ways in which he uses the resources that he controls? Do overly aggressive supervisors use their coercive powers to induce compliance more frequently than less aggressive supervisors?

In essence we are asking if there is a psychology of the use of social powers within industry? With the exception of the important and systematic research that has derived from French and Raven's (1959) classification of social powers, there has not been much interest in this question. Yet it is clear that one can distinguish between questions concerned with leadership style (i.e., how the leader behaves with respect to decision making, consideration, task orientation, etc.) and with leadership power. Several studies have reported that the influence of a leader over subordinates varied systematically with his control of resources, independently of his style of leadership (Kipnis, 1958; Pelz, 1951).

The purpose of the present study was to investigate one aspect of the relation between social power and supervisory behavior. The study investigated (a) the range of social powers available to supervisors when correct-

¹ The collection of data was made possible through the helpful cooperation of Charles A. Thomas and the American Association of Industrial Management, National Metal Trade Association.

² Requests for reprints should be sent to David Kipnis, Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.

ing subordinate behavior and (b) the situational and personal factors influencing the supervisors' use of these powers. The present research conducted in an industrial setting was a follow-up of a previously unpublished investigation (Kipnis, Lane, & Frankfurt, 1961) of naval supervisors. The military study found that a variety of personal and situational factors influenced the military supervisors' choice of corrective powers.

The following findings from the military study are relevant to the present research:

1. Military supervisors supervising large numbers of men relied upon their legal powers to punish by placing subordinates on report—a procedure often culminating in court-martial.

2. As the complexity of the problem increased, military supervisors (a) more frequently transferred subordinates to a different set of duties and (b) increased the number of corrective powers used.

3. There appeared to be a "treatment of choice" associated with problems presented by subordinates. Different powers were invoked by the supervisor according to the type of problem presented by the subordinate.

4. Experienced supervisors were more likely than inexperienced supervisors to correct directly a subordinate's behavior. Inexperienced supervisors either referred subordinates to someone else, or relied upon their legal powers. In a follow-up study on this last point (Kipnis & Lane, 1962), it was found that supervisors who lacked confidence in their leadership talents were more likely to use the latter forms of corrective powers.

It was clear that the range of corrective powers reported by these military supervisors was not chosen because of whim or individual idiosyncracies. No supervisor mentioned physical coercion, fines of money, or excessive restrictions of personal liberties. The military supervisors' descriptions of their behavior were in fact descriptions of the constraints imposed upon them by the organizational structure. If the magnitude and variety of corrective powers permitted the supervisor were increased or decreased, it could be expected that the reports of the supervisors would be correspondingly altered.

METHOD

The same procedure used in the military study was used in this investigation. An open-ended questionnaire was administered to a sample of 184 supervisors from five different companies engaged in light manufacturing. The questionnaire was given on the first day of a supervisory training course. The questionnaire asked each supervisor to describe an incident that occurred within the past year in which a subordinate's behavior was below average. In addition, supervisors were asked to describe what was done about the incident by themselves or by others. Respondents were asked to give the following information concerning their subordinates: (a) number directly supervised, (b) union or nonunion members, and (c) hourly or salaried pay. In addition, they were asked how many years they had been a supervisor. This procedure provided a listing of the kinds of corrective powers available to the supervisors, as well as the frequency with which each of these corrective actions was used.

Usable returns were obtained from 131 supervisors. Of the remainder, 25 returns described incidents that happened to someone else and 28 described cases that involved female subordinates. It was decided to analyze only the male returns. The overwhelming majority of the supervisors (89%) were directing hourly paid, blue-collar workers. Hence the findings should not be generalized to salaried, white-collar samples.

The problems and actions taken by the supervisors were coded according to a classification system used in the naval study. This system was used directly with the industrial sample, with the addition of the category of man fired and the substitution of the category written warning for the category written report.

The kinds of problems presented by subordinates were classified as follows.

1. Attitude—The subordinate showed a lack of interest in the company, work, or personal advancement.

2. Discipline—The subordinate failed to follow the rules of conduct prescribed by the company.

3. Work—The subordinate failed to maintain minimum standards in performing work.

4. Appearance—The subordinate failed to dress appropriately.

Corrective Actions

The ways in which the supervisors reported handling the problems were classified into eight categories. Since many supervisors reported taking more than one action, multiple coding was used. However, this multiple coding was used only between, and not within, categories.

1. Verbal: (a) Diagnostic talk—An attempt was made by the supervisor to find out the reasons for the subordinate's unacceptable behavior. (b) Corrective talk—The supervisor pointed out the consequences of the subordinate's substandard behavior, and/or discussed ways in which the subordinate

could improve. There was no indication that the supervisor tried to find out why the subordinate was behaving as he did.

2. Increased supervision: (a) Extra instruction—Additional or extra instruction in the area of the subordinate's poor behavior was assigned, or the supervisor spent extra time with the subordinate, closely directing his work. (b) Inspection—Frequent check-ups were made on the subordinate's performance.

3. Situational change: (a) Reassign—New or additional duties were assigned to the subordinate, or the subordinate was reassigned to a different task. The reassignment was not made for purposes of punishment. (b) Transfer—The subordinate was transferred to a different department or shift.

4. Penalty: (a) Reprimand—The subordinate was rebuked for his below-standard behavior. (b) Extra work—The supervisor assigned difficult or dirty work. (c) Reduced privileges—The subordinate was penalized by temporarily denying or reducing privileges.

5. Refer: The subordinate was referred to a superior, a peer, a specialist, or to the personnel office. The supervisor consulted with others as to what to do. Included here were two cases where the supervisor did nothing to correct performance.

6. Written warning (report for military): The subordinate was given an official written warning from the company advising him that his performance was unacceptable.

7. Man fired: The subordinate was discharged from the company.

8. Example: The supervisor acted as a model in the subordinate's problem area. This involved no direct attempts at instruction.

Outcomes

While not requested, some supervisors wrote on the questionnaire that the corrective actions taken had improved the subordinate's behavior and no further difficulty with the subordinate ensued. This information was coded in the following manner: (a) improvement—the actions taken corrected the behavior; and (b) not reported.

RESULTS

Table 1 shows the kinds of supervisory problems reported by the supervisors. For comparative purposes, the distribution of problems reported by the naval sample is also shown.

Examination of the totals in Table 1 shows that problems of appearance were mentioned more frequently by the military, while problems of discipline were mentioned less frequently. Two-thirds of the problems mentioned by both industrial and military supervisors involved getting subordinates to do their work properly. Problems of motivation, as reflected in the incidence of attitudinal problems, were mentioned by up to 18% of the industrial sample and 23% of the military sample. These latter findings illustrate the well-known fact that socioemotional problems constitute an important aspect of supervision.

Table 1 also indicated that most subordinates manifested one problem at a time to their supervisors. In 17% of the industrial descriptions and in 28% of the military descriptions, subordinates were described as manifesting two or more supervisory problems simultaneously. We shall return to this finding when we consider the relation between problem complexity and solutions attempted.

The corrective actions taken by the industrial sample are shown in Table 2. For purposes of comparison, the actions taken by military supervisors are shown also. It should be noted that the military classifications do not include the category man fired, since such a corrective action is not used in the military. Because 43% of the industrial supervisors and 51% of the military stated they used more than one corrective action, the total percentages shown in Table 2 exceed 100%.

Both industrial and military supervisors relied upon a wide variety of powers to correct performance. Many of the actions were

TABLE 1

SUBORDINATE PROBLEMS REPORTED BY SUPERVISORS

| Problem of subordinate | Industry ^a | Military ^b |
|---|-----------------------|-----------------------|
| Appearance | 0% | 9% |
| Attitude | 8% | 7% |
| Discipline | 27% | 14% |
| Work | 47% | 42% |
| Work and attitude | 8% | 7% |
| Work and discipline | 5% | 5% |
| Work and appearance | 1% | 6% |
| Other multiple combinations of problems | 3% | 10% |
| Totals | 100% | 100% |
| Totals | | |
| Total work problems mentioned | 62% | 67% |
| Total discipline problems | 36% | 24% |
| Total attitude problems | 18% | 23% |
| Total appearance problems | 2% | 23% |

^a N = 131.

^b N = 146.

based upon the supervisor's persuasive powers; others relied upon actual or verbal threats of punishment; others upon the expert knowledge of the supervisor; and still others upon the power of the supervisor to make changes in the work environment of the subordinate. Between 8–10% of both the industrial and military supervisors invoked higher administrative levels through the actual firing of the subordinate, or through the use of official warnings, or through formal reports. Finally, about 15% of both groups consulted with someone as to what to do about the problem, or referred the subordinate elsewhere. In essence, these listings represent the range of powers that the industrial and military organizations allowed their supervisors to use.

It may also be observed in Table 2 that industrial supervisors were less likely than military supervisors to attempt direct changes in their subordinates' behaviors. That is, significantly fewer industrial supervisors reported using extra instruction ($p < .01$), or changing the pattern of the subordinate's job duties in an attempt to correct performance ($p < .01$). In terms of direct punishments that did not involve formal proceedings, industrial supervisors more frequently relied upon reprimanding their subordinates ($p < .01$), whereas military supervisors used punishments that directly changed the subordinates' working conditions through extra work assignments or reduced privileges.

Relation between Problems and Actions

The first study found a "treatment of choice" associated with each problem encountered. To determine if this held in the present sample, the 109 supervisors who reported that their subordinates presented only a single problem were sorted into three problem areas of attitudes, discipline, and work. The distribution of corrective actions taken for each problem area was then determined.

Diagnostic talks were used more frequently in incidents involving attitudes or discipline than in incidents involving work (31% versus 15%, $p < .05$). Increased supervision was used more frequently in problems of work than in problems involving attitudes and

TABLE 2
CORRECTIVE ACTIONS TAKEN BY SUPERVISORS

| Action taken by supervisor | Industrial ^a | Military ^b | Comparison of industrial versus military p^c |
|----------------------------|-------------------------|-----------------------|--|
| Verbal | | | |
| Diagnostic talk | 23% | 18% | <i>ns</i> |
| Corrective talk | 42% | 33% | <i>ns</i> |
| Increased supervision | | | |
| Extra instruction | 19% | 33% | $< .01$ |
| Inspection | 7% | 10% | <i>ns</i> |
| Situational change | | | |
| Reassign | 3% | 18% | $< .01$ |
| Transfer | 8% | 1% | — ^d |
| Penalty | | | |
| Reprimand (verbal) | 16% | 5% | $< .01$ |
| Extra work | 0% | 9% | — ^d |
| Reduced privileges | 1% | 8% | — ^d |
| Refer | 15% | 15% | <i>ns</i> |
| Written warning (report) | 7% | 10% | <i>ns</i> |
| Man fired (industry only) | 8% | | |
| Set example | 1% | 7% | — ^d |

^a $N = 131$.

^b $N = 146$.

^c p values obtained through chi-square analyses in which the number of industrial and military supervisors stating they carried out the action were compared.

^d Chi-square not computed because of small N s involved.

discipline (45% versus 6%, $p < .01$). Finally 14% of the supervisors with discipline problems stated that the subordinate was fired, as compared to 0% of the supervisors with attitude problems and 3% of the supervisors with work problems. It appears that subordinates are most likely to be fired for breaking rules. These findings closely parallel the original military findings. In that study, poor work was associated with increased supervision, poor attitudes with diagnostic talks, discipline problems with official reports and/or diagnostic talks, and poor appearance with frequent inspections.

Another finding had to do with the complexity of the problems presented by the subordinate. In both studies, when the subordinate presented two or more problems simultaneously (e.g., poor attitudes and poor work), supervisors changed the job environment of the subordinate. The action of transfer was reported by 18% of the industrial supervisors with complex problems and 5% of the industrial supervisors with simple problems ($p < .10$). In the military study, the action of reassignment was used by 39% of the supervisors reporting complex problems and 10% of the supervisors with simple problems ($p < .01$).

In addition to the specific kinds of actions used, problem complexity was also related to the number of corrective actions used by the supervisor. Two or more corrective actions were used by 38% of the industrial supervisors with simple problems and by 62% of the industrial supervisors with complex problems ($p < .05$). In the military study two or more corrective actions were reported by 41% of the supervisors with simple problems and 76% of the supervisors with complex problems ($p < .01$).

Years of Experience as a Supervisor

The military study found that less experienced supervisors more frequently referred their subordinates to someone else. This finding was repeated in the present study. Twenty-seven percent of the supervisors ($N = 40$) with 2 yr. or less experience stated that they referred the subordinate's problem to someone else as compared to 7% of the industrial supervisors ($N = 30$) with 3–8 yr. of experience and 12% of the industrial supervisors ($N = 42$) with 9 or more yr. of experience ($p < .05$).³ However, the present study found no evidence that inexperienced supervisors used official warnings as was true of inexperienced military supervisors.

Number of Men Supervised and Actions Taken

There is general agreement that the more men the supervisors are required to direct, the less able they are to give their men personal attention (Dale, 1959; Yoder, 1956). Support for this contention was found in the military study, in that military supervisors directing large numbers of men were less likely to use extra instruction and more likely to place subordinates on official report.

In the present study it was also found that as the number of men supervised increased, the use of official warnings increased. Seventeen percent of the supervisors ($N = 35$) directing 15 or more subordinates, 3% of those ($N = 36$) directing 7–14 subordinates, and 0% of those ($N = 40$) directing less than 7 subordinates used official warnings as a means of correcting subordinates' performance.

³ Nineteen supervisors did not report their years of experience.

ances ($p < .01$, comparing 14 or less versus 15 or more). The use of extra instruction, however, was not related to number of men supervised. There was no relation between years experience as a supervisor and number of men supervised.⁴

Union–Nonunion

There were only 12 supervisors of union members. While this number was too small for statistical analysis, inspection revealed that in comparison to nonunion members, supervisors were less likely to talk to union members, less likely to spend time with them in extra instruction, and more likely to reprimand them or to issue official warnings as preferred means of correcting performance.

Actions and Outcomes

The more corrective actions (excluding firings) the supervisor reported, the more likely he was to state that the subordinate's performance improved. Twenty-seven percent of the supervisors who used one corrective action and 52% of the supervisors who used two or more corrective actions stated that their subordinates' behavior improved ($p < .05$).

DISCUSSION

The findings point to the important role of corrective powers in supervisory decision-making and problem-solving behaviors. It appears that the range of corrective powers controlled by supervisors represents the range of potential solutions that they may try when correcting subordinate's performance. This problem-solving interpretation was suggested in the present study by the relationship between the kind and complexity of the problem presented by subordinates and the kind and number of corrective actions used by supervisors. For example, complex problems led to the supervisor's trying more corrective actions, and each kind of subordinate problem evoked a different corrective action. It would follow from this interpretation that as the range of corrective powers that is allowed the supervisor is increased or decreased by management, one could expect corresponding in-

⁴ Twenty supervisors did not report the number of men they supervised.

creases or decreases in the supervisors' abilities to correct subordinates' performance.

It was further found that industrial supervisors were less directive in their attempts to correct behavior than military supervisors. The military supervisor more often corrected subordinates' performance by changing their duties, increasing the amount of direct supervision, and by penalizing them by assigning extra work and/or invoking penalties. The industrial supervisor relied more on his persuasive powers through use of diagnostic talks, corrective talks, or verbal reprimands.

Does this mean that the industrial supervisor is more "tenderhearted," or does not have a tradition of using more direct forms of action? Even a cursory reading about the industrial management scene from the 1880s through the beginning of World War II would indicate that this is not the case. Prior to World War II the industrial supervisor was more likely to be the person allowed by management to have a major voice in hiring, firing, demotions, layoffs, wages, and the general regulating of the working conditions of his subordinates. A supervisor of that time was nicknamed aptly "bull of the woods." Since that time, however, supervisory powers have been reduced by union contracts, delegation of responsibilities to staff personnel, and broad changes in management philosophy. As a result of these recent events, industrial supervisors control a smaller range of corrective powers than do their military counterparts. This reduced control is believed reflected in the greater reliance of industrial supervisors, in the present sample, upon verbal persuasion rather than upon more direct attempts to influence subordinates.

Further questions concerned with leadership powers can be organized into three areas. The first is concerned with situational factors that influence the use of supervisory powers. The present study found differences between military and industrial organizations. Within each organization, span of control and kind of problem also influenced supervisor's corrective actions.

A second question is concerned with the supervisor's own response to his possession of social powers. An interesting study by Lange and Jacobs (1960) of the actual day-

to-day behaviors that distinguish between effective and ineffective platoon leaders revealed that ineffective leaders used their powers to reward and punish inappropriately. Instead of rewarding or punishing subordinates in terms of actual job performance, they used these sanctions to curry favor or to punish irrationally. Our research (Kipnis & Lane, 1962) strongly suggests that inexperience and lack of confidence may make the supervisor reluctant to use the full range of powers that he controls.

The third question that requires attention has to do with the subordinate's response to supervisory powers. Does reprimanding an employee do any good besides allowing the supervisor to "blow off steam"? Does reassigning the problem employee improve his performance? In what ways are the employee's self-esteem, ideological allegiances, and morale affected by reliance upon the various forms of power? Many students suggest that the possession of a broad complex of powers by a leader causes the subordinate to feel uneasy, distrustful, and reluctant to reveal weaknesses in himself to his supervisors (Hutchins & Fiedler, 1960; Mulder, 1959; Wispé & Lloyd, 1955). These subordinate feelings in turn lead to distortions in upward communications and approach-avoidance conflicts over interactions with supervisors (Mellinger, 1956; Read, 1962). Thus it may prove that increasing the range of powers controlled by the supervisor will improve his problem-solving abilities, but at the expense of provoking more guarded and defensive behaviors among subordinates.

REFERENCES

- BALES, R. F. The equilibrium problem in small groups. In T. Parsons, R. F. Bales, & E. A. Shills (Eds.), *Working papers in the theory of action*. Glencoe, Ill.: Free Press, 1953.
- CARTER, L. F. *Leadership and small group behavior*. Paper presented at the Second Conference on Social Psychology, University of Oklahoma, 1952.
- DALE, E. *Planning and developing the company organization structure*. Research Report No. 20. Washington, D. C.: American Management Association, 1959.
- FIEDLER, F. E. The contingency model: A theory of leadership effectiveness. In H. Proshansky & B. Seidenberg (Eds.), *Basic studies in social psychology*. New York: Holt, Rinehart & Winston, 1965.

- FLEISHMAN, E. A. The description of supervisory behavior. *Journal of Applied Psychology*, 1953, 37, 1-6.
- FRENCH, J. R. P., & RAVEN, B. The bases of social power. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor, Mich.: Research Center for Group Dynamics, 1959.
- HUTCHINS, E. B., & FIEDLER, F. E. Task oriented and quasi-therapeutic role functions of the leader in small military groups. *Sociometry*, 1960, 23, 393-406.
- KAHN, R. L., & KATZ, D. Leadership practices in relation to productivity and morale. In D. Cartwright & A. Zander (Eds.), *Group dynamics*. (2nd ed.) Evanston, Ill.: Row, Peterson, 1960.
- KIPNIS, D. The effects of leadership style and leadership power upon the inducement of an attitude change. *Journal of Abnormal and Social Psychology*, 1958, 58, 173-180.
- KIPNIS, D., LANE, W. P., & FRANKFURT, L. *Leadership problems and practices of petty officers*. Washington, D. C.: Naval Personnel Research Activity Research Report 61-30, 1961.
- KIPNIS, D., & LANE, W. P. Self-confidence and leadership. *Journal of Applied Psychology*, 1962, 46, 291-295.
- LANGE, C. J., & JACOBS, T. O. *Leadership in army infantry platoons: Study II*. U. S. Army Infantry Research Unit, Fort Benning, Georgia, 1960.
- MELLINGER, G. D. Interpersonal trust as a factor in communication. *Journal of Abnormal and Social Psychology*, 1956, 52, 304-309.
- MULDER, M. Power and satisfaction in task-oriented groups. *Acta Psychologica*, 1959, 16, 178-225.
- PELZ, D. C. Leadership within a hierarchical organization. *Journal of Social Issues*, 1951, 7, 49-55.
- READ, W. H. Upward communication in industrial hierarchies. *Human Relations*, 1962, 15, 3-16.
- WISPE, L. G., & LLOYD, K. E. Some situational and psychological determinants of the desire for structured interpersonal relations. *Journal of Abnormal and Social Psychology*, 1955, 51, 57-60.
- YODER, D. *Personnel management and industrial relations*. New York: Prentice-Hall, 1956.

(Received December 5, 1968)

IMPACT OF EMPLOYEE PARTICIPATION IN THE DEVELOPMENT OF PAY INCENTIVE PLANS:

A FIELD EXPERIMENT

EDWARD E. LAWLER III¹ AND J. RICHARD HACKMAN

Department of Administrative Sciences, Yale University

A study of effects of employee participation in the development of pay incentive plans. The *Ss* were part-time workers who clean buildings in the evenings. Three autonomous work groups developed their own pay incentive plans to reward good attendance on the job (Condition A). These plans were then imposed by the company on other work groups (Condition B). There were two groups of control *Ss*: One talked with *Es* about job attendance problems but received no additional experimental treatment, and the other received no treatment. A significant increase in attendance followed only Condition A. Possible reasons cited: (a) participation caused *Ss* to be more committed to the plan; (b) *Ss* who participated in the development of their plan were more knowledgeable about it; and (c) participation increased the employees' trust of the good intentions of management with respect to the plan.

Literally thousands of different pay incentive plans have been developed and used in an effort to increase the motivation of employees in work organizations. These plans have tried to motivate a number of different behaviors: productivity, sales, cost reduction, job attendance, etc. They also have differed widely in form: Some have used individual incentives, others have provided rewards on a group or organization-wide basis; some have been based on small units of behavior while others have been based on relatively long-term performance.

A good deal of research has attempted to determine the relative effectiveness of different kinds of incentive plans. For example, group plans have been compared with individual plans and bonus plans have been compared with salary increase plans. The results suggest that these characteristics of plans do affect their success (see, e.g., Opshal & Dunnette, 1966; Rothe, 1960; Viteles, 1953). Nevertheless, it is striking to note the number of instances in which an identical plan is successful in one situation but unsuccessful in

another (e.g., Whyte, 1955). Apparently the success of pay incentive plans is determined by more than just the operating mechanics of the plans themselves.

Virtually no research has been done to identify the nonmechanistic factors that are important in determining the success of pay plans. The present study attempts to determine the impact of one such nonmechanistic factor on the success of a pay plan. Specifically, this study examines how the way a pay plan is developed and implemented influences its effectiveness.

Previous research (e.g., Coch & French, 1948) suggests several reasons why the way a plan is developed and introduced may be crucial in determining how successful it is.

If a pay plan is to be successful it is important that all the participants understand it, be committed to it, and believe that it will be administered fairly. Clearly, all of these factors can be influenced strongly by the way the plan is developed and introduced. A pay plan that is developed by a mistrusted management and imposed upon workers is not likely to be understood by employees, nor are the workers likely to be committed to its success. On the other hand, a plan that is participatively developed by workers is very likely to be understood by them and they are more likely to be committed to its success. The plan also is likely to be appropriate to the

¹ The authors would like to thank H. Elston, M. Nunes, and R. Breck for their cooperation with the study. Wendy Silin deserves special mention for her help with the data analysis.

Requests for reprints should be sent to Edward E. Lawler III, Department of Administrative Sciences, Yale University, 2 Hillhouse Avenue, New Haven, Connecticut 06520.

situation in which the workers find themselves. The basic hypothesis of the study therefore is: Pay incentive programs will be more effective if they are participatively developed than if they are imposed upon a group of employees by management. It should be noted here that there is no intention to indicate that just because a pay plan is participatively developed it will be successful. It seems unlikely that any plan will work if it is not set up properly and administered well, even if it is participatively developed.

METHOD

Research Strategy

The basic hypothesis of the study was tested in a field experiment. The experimental approach was chosen because it allows the causal impact of the experimental factor (in this case, participative development of the pay plan) to be assessed with relatively little ambiguity. The study was done in a field setting to ensure that the manipulation would be realistic and important to Ss and to increase the likelihood that the results would be generalizable to other field settings.

Research Site and Subjects

The research was conducted in a small company that provides building maintenance services on a contract basis. The Ss were part-time employees of the company, who clean buildings during the evening. Most Ss worked 4 hr. a night. Prior to the study, the company had experienced extremely high rates of absenteeism and turnover among these employees.

The Ss worked in groups ranging in size from 2 to 25. There were about 15 such groups in the company at the time of the study. Each group was responsible for doing all the cleaning work in one building. Although the groups did similar work, they were highly autonomous. The employees always reported for work at the building they were to clean and never came to the company offices. Because of this there was virtually no contact between employees in different work groups.

The Ss tended to have very low educational levels, and most were members of minority groups. A number of them were illiterate. Approximately half the Ss were women, many of whom were housewives during the day. For most of the male Ss, the maintenance work was a second job. The Ss ranged in age from 16 to over 70.

Procedure

Nine work groups were involved in the experiment. Three designed their own incentive plans (the participative groups), two had incentive plans imposed on them (the imposed groups), two talked with the researchers but their pay plans were not changed,

and two received no treatment at all (the control groups). There were no apparent differences among the groups assigned to the different treatment conditions. The groups all worked in comparable buildings, and the members of the different groups were similar demographically (e.g., age, education, experience, and social class).

Participative groups. Three work groups (of 10, 9, and 8 members) were selected as participative experimental groups. Both of the authors met with one of the groups to help the employees develop an incentive plan, and each of the authors met separately with one of the other two groups. Both researchers worked with one of the groups, so that they would be able to behave in similar ways with the groups they were handling on an individual basis.

In all cases the researchers were introduced to the employees during regular working hours by a member of top management. The manager told the employees that the company was concerned about high rates of absenteeism, and expressed his hope that the employees would work with the researchers in developing an appropriate plan for rewarding good attendance.

The manager then left and the researcher opened the discussion by emphasizing that it was his objective to help the employees develop a plan and not to tell them what kind of plan they should develop. In all three groups an extensive discussion followed this introduction. During the initial phases of the discussion the workers expressed a great deal of mistrust of the researcher and they displayed considerable hostility toward both the researcher and the company. They continually demanded to know what kind of plan the company wanted them to develop, and they asked why the researcher was interested in working with them. The researcher allowed the initial discussion to continue for about 45 min., at which point he asked the employees to talk things over among themselves and said that he would be back the next night to continue the discussion. The employees then returned to their work and the researcher left.

For all three groups the second meeting resulted in much more progress than did the first. Although still suspicious of the motives of the company and the researcher, the employees began to discuss what might constitute an acceptable plan. Much of this discussion focused on how large the bonus should be. During the discussion the researcher took on the role of a resource person for the group. At no time did he suggest a plan, although if he was asked about a specific idea he did react to it by stating a few general principles about what makes for successful pay incentive plans (e.g., it is important to relate pay to behavior). One group decided to ask for a very large amount ("since the company will cut whatever we ask for in half anyway"); the other two groups seemed more responsible and settled on dollar amounts that were less than the company had originally anticipated offering.

By the third meeting in one group and the fourth in the others, a plan had been developed and agreed to by all group members. The three plans that were

developed by the groups did show some important differences. Two groups wanted their bonus to be computed on a weekly basis, while the other group wanted a monthly bonus. There were also differences in the size of bonuses requested and in the number of days of sick leave that should be allowed. The researchers presented the plans to management, and they all were quickly accepted with minor alterations. The alterations involved adjustment of the amount of the bonuses so that they would be equivalent for all three groups and specification of what would constitute an "excused" absence from work.

As finally instituted, all plans offered cash bonuses of about \$2.50 per week for perfect attendance. In one plan the bonuses were to be computed and paid at the end of each month, while the other two plans were on a weekly basis. When the plans were instituted, a manager of the company returned to each group to answer any final questions and to explain why the changes had been made in the original proposals of the work groups. He did not ask them to approve the changes formally.

Imposed groups. Plans identical to those developed by the employees in the participative condition were imposed by the company on two other groups ($N=13$; $N=26$). One group received the weekly plan and the other received the monthly plan.

The same manager who had worked with the researchers in the participative condition instituted the plans in these groups. Accompanied by one of the researchers, he met with the groups and explained why a bonus plan was being instituted and how it was to operate. He spent considerable time with each

of the groups, and appeared to do an adequate job explaining the plans and showing the employees how they personally would benefit if they came to work more regularly.

Control groups. The researchers visited two other groups ($N=9$; $N=8$) and talked with these employees at length about incentive plans and about problems of absenteeism and turnover. It was stressed in these meetings that the researchers were interested in studying how people react to wages and that the company was concerned about the current high rates of absenteeism. No changes were made in the pay plans of these two groups. In two other groups ($N=26$; $N=8$) no changes in the pay plans took place and the researchers did not meet with the employees, but the attendance of the groups was monitored.

RESULTS

Results are presented in Figures 1 and 2. The data are expressed in terms of the percentage of an employee's scheduled work week that he actually worked. For most employees the work week was 20 hr. long. Thus, if an employee worked 10 hr., he was scored as working 50% of his scheduled hours; if he worked 15 hr., he was scored as 75%. Figure 1 shows the mean percentage of scheduled hours actually worked for Ss in the three participative groups. Figure 2 shows analogous data

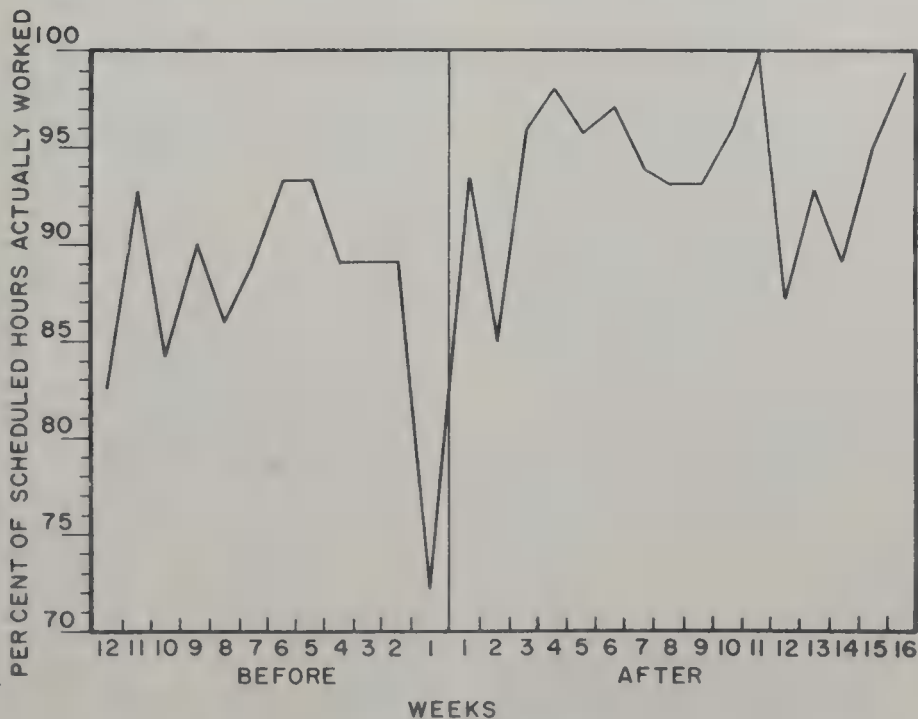


FIG. 1. Mean attendance of the participative groups for the 12 wk. before the incentive plan and the 16 wk. after the plan. (Attendance is expressed in terms of the percentage of hours scheduled to be worked that were actually worked.)

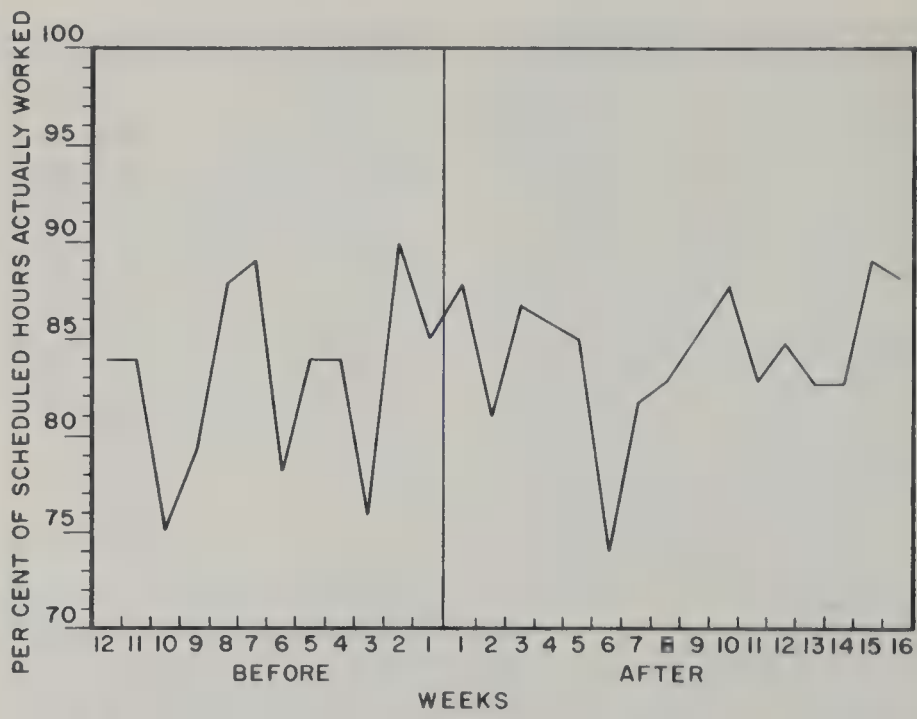


FIG. 2. Mean attendance of the imposed groups for the 12 wk. before the incentive plan and the 16 wk. after the plan. (Attendance is expressed in terms of the percentage of the hours scheduled to be worked that were actually worked.)

for Ss in the two groups that had incentive plans imposed on them. In both figures, data are presented for the 12 wk. before the plans were instituted and for the first 16 wk. after the plans went into effect.

Before the incentive plans were introduced, the average employee in the participative groups worked 88% of his scheduled hours; after the plan went into effect, the average employee worked 94% of his scheduled hours. This before-after difference was tested for statistical significance by a median test yielding a chi-square of 9.35 ($p < .001$).²

As is shown in Figure 2, there was no improvement in attendance for groups in which the identical incentive plans were imposed by management. Before the imposed plans were instituted, the average employee worked 83% of his scheduled hours; in the 16 wk. after the plans were put into effect, the figure remained at 83%.

Data gathered from the control groups

² This before-and-after comparison essentially uses each participative group as its own control group. This was done because the initial differences between the attendance levels of the imposed and the participative groups make post-comparisons between these groups artificial.

(whose pay plans were unchanged) showed no significant changes during the period of the study.

Thus, the data show that employee attendance improved only in those groups that participatively developed their own incentive plans. Neither the incentive plan alone nor participation and discussion alone yielded any changes in attendance.

DISCUSSION

The results of this study strongly support the notion that attention to the technical characteristics of a pay plan alone (i.e., the mechanics of its design and administration) may be insufficient to ensure the success of the plan. Indeed, the data suggest that participation in the development and implementation of a plan may have more of an impact on the effectiveness of a plan than the mechanics of the plan itself.

Why should participation be so important to the success of a pay incentive plan? One possibility is that participation can improve the quality of decisions that are made (Vroom, 1964). Thus, it could be argued that the participatively developed plans in this

study were uniquely suited to the groups that developed them—but somehow inappropriate for the other groups on which the plans were imposed. This explanation seems unlikely in the present study, however, since all the work groups in the company were highly similar and there was nothing especially unique about the plans developed by the three participative groups.

It did seem that members of the participative groups more fully understood the plans than did members of the imposed groups. Despite a carefully rehearsed introductory talk given by the company manager, the imposed groups did not receive as much information about the plan as did the participative groups. They did not feel as free to ask questions about the plan as did the participative groups, and they did not have as much time to think about questions or to ask them as did the participative groups. The imposed groups received all their information about the plan in one session, whereas members of each participative group talked together about the plan for several hours over a week or two—possibly increasing their understanding of the plan and its implications.

It also appeared that the participative groups were more committed to the success of the plans than were the imposed groups. There was evidence that the plans were viewed as “just another attempt by management to exploit us” by some members of the imposed groups. By participating in the development of the plans, many members of the participative groups appeared to become more trusting of management’s intentions to administer the plans fairly. Their pride in “owning” the plans, coupled with the increased trust of management, may have enhanced considerably the desire of the participative employees to cooperate in making the plans a success.

A final comment may be in order about the characteristics of the *S* population studied. The employees were, almost without exception, of a low socioeconomic class, and all were working at low-level jobs. Thus, it is perhaps surprising that they responded to the opportunity for participation as well as they did. For most of them it was the first time they had ever had an opportunity to contribute meaningfully to any decision making about their jobs. There was hostility and suspicion at the outset of the experiment. Yet, after the initial discussion with the researchers, a substantial number of the employees began to respond to the challenge of developing a viable incentive plan—and they ultimately came up with plans that would have to be considered technically adequate. Thus it appears that—if a researcher or a manager is willing to deal with some initial hostility and suspicion—it should be possible to involve most employees in meaningful decision making about their jobs. And, if the results of this study have generalizability, the payoff for both the employees and the organization should make the effort well worthwhile.

REFERENCES

- COCH, L., & FRENCH, J. R. P., JR. Overcoming resistance to change. *Human Relations*, 1948, 1, 512-532.
- OPSHAL, R. L., & DUNNETTE, M. D. The role of financial compensation in industrial motivation. *Psychological Bulletin*, 1966, 66, 94-118.
- ROTHER, H. F. Does higher pay bring higher production? *Personnel*, 1960, 37, 20-27.
- VITELES, M. S. *Motivation and morale in industry*. New York: Norton, 1953.
- VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.
- WHYTE, W. F. *Money and motivation: An analysis of incentives in industry*. New York: Harper, 1955.

(Received December 6, 1968)

CONSUMER ATTITUDES TOWARD AUTO VERSUS PUBLIC TRANSPORT ALTERNATIVES¹

FRANK T. PAINE, ALLAN N. NASH,² STANLEY J. HILLE,

Department of Business Administration, University of Maryland

AND G. ALLEN BRUNNER

College of Business Administration, University of Toledo

Two pilot studies of consumer attitudes toward auto versus public transport modes were conducted in Baltimore, Maryland, and Philadelphia, Pennsylvania ($N = 550$ and 471 , respectively), using a quasi self-administered Likert-scaled questionnaire. Results indicate a consistently favorable preference for auto over a wide range of mode attributes but with marked differences in magnitude of preference between attributes. Significantly different patterns were also found between inner-city and suburban preferences. Implications for changing mode-use patterns are discussed briefly.

Most transportation consumer research has been of the origin-destination variety that provides a detailed description of the traveler, mode used, and trip purpose (Gilat, 1963). Questions have been answered from this research about where and how people traveled, but an explanation of their behavior generally has not evolved. A few studies, however, have partially focused on consumer attitude measurement emphasizing the identification and assessment of consumer values relevant to transport selection decisions (Ackoff, 1965; Lansing, Mueller, & Barth, 1964; Mahoney, 1964; Stanford Research Institute, 1965).

Although in most cases these efforts have achieved stated objectives, many have had several limitations that restricted the generalization of their results. One of the most severe has been the small selected samples used. Another has been the narrowness of focus in terms of such variables as mode, trip, and/or characteristics of users. The latter

made it difficult to compare and contrast results between studies because different variables were included in each. In some cases the method of collecting data was not carefully constructed and/or evaluated. Finally, the designs were based on the proposition that the researchers knew which modal characteristics to study and how to define them. Usually abstract variables such as "convenience, comfort, status, congestion, flexibility, expense, etc." have been used as inputs without careful specific definitions. Research on various rating scales suggests that trait scales with fairly specific behavioral descriptions are more reliable and valid than scales evaluating "how much" of a trait is possessed by a ratee with abstract scale gradations (Stockford & Bissell, 1949). Experience with checklists also shows significant improvement in measurement compared to results obtained when global traits are used (Miner, 1969).

Two pilot studies described herein attempted to alleviate partially some of the above limitations.

First, an attempt was made to provide a more comprehensive coverage of significant variables affecting modal choice decisions. Previous models had only a modicum of success for predicting modal split decisions, probably because these decisions are more complex than they were originally thought to be. As few as two variables have been used (travel time and cost) to predict modal choice, and many studies include from four to six

¹ This study was supported by United States Department of Commerce, Bureau of Public Roads, Contract CPR-11-0960, Project Director: A. N. Nash, and is partially based on G. A. Brunner, S. J. Hille, A. N. Nash, F. T. Paine, R. E. Schellenberger, and G. M. Smerk, *User Determined Attributes of Ideal Transportation System*, College Park: University of Maryland, 1966, 228 pp., and F. T. Paine, A. N. Nash, S. J. Hille, and G. A. Brunner, *Consumer Conceived Attributes of Transportation*, College Park: University of Maryland, 1967, 177 pp.

² Requests for reprints should be sent to Allan N. Nash, Department of Business Administration, University of Maryland, College Park, Maryland 20742.

variables. Their results suggested that the development of a valid prediction model for modal choice decisions requires the incorporation of several attributes into the prediction milieu, and requires that the model be sensitive to the complex interrelationships existing among attributes.

Second, the attitude instrument developed sought to determine both the importance of and satisfaction with modal attributes. The importance of a particular attribute is probably a function of both the underlying strength of the human need or needs to which it is related and its present satisfaction level. The inclusion of satisfaction items with the importance of items sought to clarify the extent to which importance of an attribute is a function of its present level of satisfaction as well as to assess overall satisfaction with existing alternative private and public modes. Thus, insight was obtained as to why people chose to travel in the modes presently used (preponderantly auto).

Third, these studies focused on the development of factor definitions by subjecting a comprehensive pool of specific items tapping particular travel characteristics and behavior to factor analysis. It was hoped that progress would result toward a definition and classification of the attributes perceived by transport users as being relatively distinct and important variables in the determination of their travel behavior.

METHOD

The first pilot study conducted in Baltimore, Maryland, involved the development and evaluation of an instrument directed at three questions of a five-question general design. These were (1) What attributes do consumers regard as salient in typical recent trips? (2) What is the relative importance of the attributes for particular trip purposes, and conglomerately? (3) To what extent, and how, are demographic characteristics of respondents related to the importance of trip mode attributes?

The second study was completed in Philadelphia, Pennsylvania, and includes specific questions relevant for all five questions in the design. The last two general questions were (4) To what extent do consumers perceive themselves as being satisfied with the attributes of auto versus public transport modes? (5) To what extent, and how, are demographic characteristics of respondents related to perceived satisfaction of trip mode attributes?

The application of a revised instrument in Philadelphia provided feedback on consistency of answers

to the original three questions under different transportation circumstances from those existing in Baltimore.

Sample

Sixty clusters (sampling points) in the Baltimore Standard Metropolitan Statistical Area and 80 clusters in the Philadelphia Standard Metropolitan Statistical Area were selected following a multistage area probability sampling design. The Baltimore sample of 350 households resulted in the completion of 550 individual questionnaires. The Philadelphia sample of 361 separate households resulted in 471 usable questionnaires.

The composition of the samples along selected social and economic characteristics (age, sex, education, head or nonhead of household, race, income, house ownership, and number in household) was compared with 1960 census data to determine representativeness. The proportion of females and the well educated appeared overrepresented in both Baltimore and Philadelphia results.

A quasi quota sampling procedure was invoked during the Philadelphia study after the tactic of having the interviewer ask for the man in the house failed to adequately correct an imbalance of too many females. This procedure improved the final balance to 58% female, 42% male. Considering the probable changes in the population during the interval between the 1960 census and the gathering of sample data, the distributions appeared to be representative for other characteristics.

Questionnaire

A questionnaire consisting of three parts and a household information cover sheet was used to collect data for the Philadelphia study.³ As indicated, this was a modified version of the Baltimore questionnaire with the addition of satisfaction questions. Part A contained a set of questions designed to elicit descriptive information about the two trip purposes asked about in Parts B and C of the questionnaire, that is, (a) the respondent's last common or usual trip to work or school, and (b) his last common or usual in-town, nonwork trip. Part B included a set of 35 items measuring the importance of attributes contained in the items along a 7-degree Likert-type interval scale, ranging from "not at all important" to "of greatest importance." These items were designed to measure exhaustively factors salient to modal choice decisions, suggested by a search of the literature and results obtained in the Baltimore study.

Part C contained a set of 33 items constructed to determine satisfaction with the Part B attributes for auto and the respondent's most likely form of public transportation for both trip purposes.⁴

³ There were two questionnaires. The questions remained the same but their order was varied to control halo and positional effects.

⁴ Two attributes were eliminated in Part C because they were inapplicable when posed as satisfaction items.

TABLE 1
PERCENTAGE OF SATISFIED RESPONSES TO EACH ITEM FOR AUTO VERSUS PUBLIC MODES
WITH ITEMS RANKED BY MEAN IMPORTANCE: WORK TRIP, TRIP PURPOSE 1

| Item no. | Description | Importance rank | % satisfied with auto | % satisfied with public transit | Difference |
|----------|--------------------------------------|-----------------|-----------------------|---------------------------------|------------|
| 28 | Arrive without accident | 1 | 89 | 86 | 3 |
| 11 | Arrive at intended time | 2 | 94 | 63 | 31 |
| 18 | Safest vehicle | 3 | 94 | 91 | 3 |
| 33 | Avoid stopping for repairs | 4 | 94 | 92 | 2 |
| 12 | Shortest distance | 5 | 97 | 64 | 33 |
| 24 | Fast as possible | 6 | 95 | 70 | 25 |
| 17 | Avoid changing vehicle | 7 | 98 | 69 | 29 |
| 2 | Vehicle unaffected by weather | 8 | 90 | 84 | 6 |
| 5 | Protected from weather while waiting | 9 | 96 | 54 | 42 |
| 1 | Shortest time | 10 | 92 | 65 | 27 |
| 26 | Avoid waiting more than 5 minutes | 11 | 98 | 55 | 43 |
| 22 | One-way cost of 25¢ rather than 50¢ | 12 | 88 | 69 | 19 |
| 27 | Comfortable | 13 | 97 | 75 | 22 |
| 13 | One-way cost of 25¢ rather than 35¢ | 14 | 88 | 67 | 21 |
| 10 | Clean vehicle | 15 | 92 | 73 | 19 |
| 14 | Feel independent | 16 | 89 | 58 | 31 |
| 16 | Avoid walking more than a block | 17 | 97 | 65 | 32 |
| 29 | One-way cost of 3¢ rather than 15¢ | 18 | 84 | 67 | 17 |
| 3 | Cost | 19 | 91 | 62 | 29 |
| 30 | Avoid unfamiliar area | 20 | 93 | 78 | 15 |
| 7 | Travel when traffic is light | 21 | 73 | 64 | 9 |
| 6 | Uncrowded vehicle | 22 | 94 | 51 | 43 |
| 19 | Package and baggage space | 23 | 94 | 57 | 37 |
| 31 | Pride in vehicle | 24 | 88 | 60 | 28 |
| 20 | New modern vehicle | 25 | 86 | 68 | 18 |
| 21 | Friendly people | 26 | 91 | 67 | 24 |
| 23 | People you like | 27 | 93 | 66 | 27 |
| 25 | Need not pay daily | 28 | 82 | 61 | 21 |
| 32 | Avoid riding with strangers | 29 | 89 | 66 | 23 |
| 4 | Listen to radio | 30 | 84 | 44 | 40 |
| 9 | Ride with people who chat | 31 | 84 | 62 | 22 |
| 15 | Look at scenery | 32 | 74 | 72 | 2 |
| 8 | Take along family and friends | 33 | 87 | 57 | 30 |

Note.—Satisfied = summated responses in Item Categories 5, 6, and 7 (i.e., generally, very well, completely satisfied, respectively).

This questionnaire was designed to be self-administered, although the interviewer was available for any needed help.

Analysis

Analytical techniques used in both studies were quite similar, so the Philadelphia analysis will be discussed. The statistical analysis for Parts B and C of the questionnaire followed procedures typically employed for attitude data. A frequency distribution of responses to each item was developed and converted into percentages for each trip purpose. The mean and standard deviation were also computed for each item.

Additionally, intercorrelations among the importance items for each trip purpose were computed and factor analyzed. The factors were rotated using the

Kaiser Varimax method.⁵ This approach was followed also for the Satisfaction section of the questionnaire; that is, factors were derived independently for the following trip purpose-mode use pairs: (a) Work Trip—Auto, (b) Work Trip—Public Transport, (c) Nonwork Trip—Auto, and (d) Nonwork Trip—Public Transport. The factors were defined using a factor loading cutoff point of .30. The relative importance and satisfaction of the factors identified in the factor analysis were derived by averaging the mean response of each component item in the dimension after other weighted alternatives were examined.

⁵ The authors wish to thank Emil Heerman, formerly of the Department of Psychology, University of Maryland, and now at the University of Nebraska, for his assistance in the selection of a factor analysis method and interpretation of the factors.

Because hundreds of possible relationships existed in the demographic and trip characteristic data, a method was devised to screen those relationships having both statistical and probable practical significance. The method involved dichotomizing or trichotomizing all responses for all demographic or trip characteristics. This resulted in 2×3 or 3×3 cell matrices. A decision rule was established which stated that only those items that had at least a 10% difference in response frequencies between at least two cells in a 3×3 or 2×3 condensed distribution when the total number of respondents is greater than 300 (or 20% when less than 300) would be considered of practical significance. The standard error of proportion was computed for these distributions and such differences are beyond the .01 level of significance.

RESULTS

Tables 1 and 2 show the relative levels of satisfaction by item for auto versus public transport for the work trip and nonwork trip. In addition, items are arrayed by average importance (as indicated by item importance means) so that importance and satisfaction levels can be determined easily for each item.

Factor analysis of the Philadelphia results indicated that about 70% of the common variance can be accounted for by eight relatively independent factors defined by similar items for both trip purposes. The discussion that follows is organized around these factors

TABLE 2

PERCENTAGE OF SATISFIED RESPONSES TO EACH ITEM FOR AUTO VERSUS PUBLIC MODES WITH ITEMS RANKED BY MEAN IMPORTANCE: NONWORK TRIP, TRIP PURPOSE 2

| Item no. | Description | Importance rank | % satisfied with auto | % satisfied with public transit | Difference |
|----------|--------------------------------------|-----------------|-----------------------|---------------------------------|------------|
| 28 | Arrive without accident | 1 | 86 | 77 | 9 |
| 18 | Safest vehicle | 2 | 89 | 81 | 8 |
| 33 | Avoid stopping for repairs | 3 | 89 | 81 | 8 |
| 5 | Protected from weather while waiting | 4 | 95 | 47 | 48 |
| 17 | Avoid changing vehicle | 5 | 96 | 68 | 28 |
| 2 | Vehicle unaffected by weather | 6 | 90 | 81 | 9 |
| 22 | One-way cost of 25¢ rather than 50¢ | 7 | 86 | 65 | 21 |
| 11 | Arrive at intended time | 8 | 93 | 66 | 27 |
| 27 | Comfortable | 9 | 97 | 69 | 28 |
| 16 | Avoid walking more than a block | 10 | 95 | 57 | 38 |
| 10 | Clean vehicle | 11 | 91 | 63 | 28 |
| 13 | One-way cost of 25¢ rather than 35¢ | 12 | 85 | 63 | 22 |
| 29 | One-way cost of 3¢ rather than 15¢ | 13 | 82 | 60 | 22 |
| 26 | Avoid waiting more than 5 minutes | 14 | 93 | 45 | 48 |
| 19 | Package and baggage space | 15 | 95 | 52 | 43 |
| 12 | Shortest distance | 16 | 91 | 55 | 36 |
| 3 | Cost | 17 | 89 | 54 | 35 |
| 24 | Fast as possible | 18 | 90 | 64 | 26 |
| 14 | Feel independent | 19 | 84 | 51 | 33 |
| 6 | Uncrowded vehicle | 20 | 91 | 48 | 43 |
| 30 | Avoid unfamiliar area | 21 | 91 | 72 | 19 |
| 7 | Travel when traffic is light | 22 | 67 | 55 | 12 |
| 1 | Shortest time | 23 | 91 | 55 | 36 |
| 21 | Friendly people | 24 | 92 | 60 | 32 |
| 23 | People you like | 25 | 91 | 60 | 31 |
| 8 | Take along family and friends | 26 | 86 | 59 | 27 |
| 20 | New modern vehicle | 27 | 84 | 62 | 22 |
| 31 | Pride in vehicle | 28 | 89 | 53 | 36 |
| 9 | Ride with people who chat | 29 | 82 | 56 | 26 |
| 15 | Look at scenery | 30 | 78 | 76 | 2 |
| 32 | Avoid riding with strangers | 31 | 91 | 58 | 33 |
| 25 | Need not pay daily | 32 | 76 | 51 | 25 |
| 4 | Listen to radio | 33 | 79 | 32 | 47 |

Note.—Satisfied = summated responses in Item Categories 5, 6, and 7 (i.e., generally, very well, completely satisfied, respectively).

TABLE 3
SUMMARY OF ABSOLUTE AND RELATIVE IMPORTANCE
AND SATISFACTION OF FACTORS COMPARING AUTO
AND PUBLIC TRANSPORT FOR THE WORK TRIP

| Factor | Mean factor importance scores ^a | Mean factor satisfaction scores ^b | | |
|------------------|--|--|--------|------------|
| | | Auto | Public | Difference |
| Reliability | 6.39 | 6.15 | 5.89 | .26 |
| Travel time | 6.14 | 6.23 | 4.99 | 1.24 |
| Weather | 5.99 | 6.18 | 5.01 | 1.17 |
| Cost | 5.50 | 5.69 | 4.97 | .72 |
| State of vehicle | 5.13 | 5.95 | 5.10 | .85 |
| Unfamiliarity | 4.62 | 6.15 | 5.38 | .77 |
| Self-esteem | 4.61 | 5.90 | 4.49 | 1.40 |
| Diversions | 4.01 | 5.86 | 4.90 | .96 |

Note.—For purposes of comparison, only those items that had high factor loadings and were common to both trip purposes were used to calculate the factor scores.
^a Listed in order of importance.
^b Highest possible score = 7.00.

in approximate order of importance. Furthermore, some of the more salient differences in perceived satisfaction with auto and public transit associated with demographic characteristics of respondents are examined. Tables 3 and 4 summarize the importance of, and satisfaction with, auto and public transport considering each factor.

Reliability of Destination Achievement

An overwhelming percentage of respondents believed the items “arrive without accident,” “avoid stopping for repairs,” and “safest vehicle” were very important. Examination of Tables 1 and 2 indicates that public transportation was rated nearly as satisfactory as the automobile for the reliability items (Nos. 28, 18, 33). These findings support the contention that although reliability and safety are important to transport users, possibilities of altering use patterns through persuasion campaigns based on a reliability and safety theme are not good.

Travel Time

The travel time factor including “arriving at the intended time” in the “shortest distance,” as “fast as possible,” and in the “shortest time” is second in importance for the work trip but less significant for the non-work trip.

Satisfaction with the auto for the work trip is quite high (average for items in factor = 6.23) and public transportation moderate (4.99). Considering both trip purposes, the differences in satisfaction (1.35 and 1.24) are among the largest between auto and public transport for this important factor, and thus travel time is probably a key determinant in the choice of auto over public transport.

Furthermore, examination of demographic variability indicates the respondents more satisfied with the auto for travel time considerations have middle to high incomes, are white, in lower age categories, and live in a single-unit dwelling. For example, Table 5 shows the percentage satisfied with the auto categorized by income, age, race, and type of dwelling for one of the travel time items (get there in the shortest time). Middle- and upper-class relative satisfaction with auto for time indicates changes to public transit will be difficult to accomplish unless the perceived time preference for auto is reduced or overcome.

Weather Factor

A weather factor follows travel time in importance and includes “vehicle unaffected by weather” and “protected from weather while waiting.” The auto is perceived to be mark-

TABLE 4
SUMMARY OF ABSOLUTE AND RELATIVE IMPORTANCE
AND SATISFACTION OF FACTORS COMPARING AUTO
AND PUBLIC TRANSPORT FOR THE
NONWORK TRIP

| Factor | Mean factor importance scores ^a | Mean factor satisfaction scores ^b | | |
|------------------|--|--|--------|------------|
| | | Auto | Public | Difference |
| Reliability | 6.34 | 5.99 | 5.64 | .35 |
| Weather | 5.98 | 6.13 | 4.84 | 1.29 |
| Convenience | 5.78 | 6.38 | 4.86 | 1.52 |
| Cost | 5.52 | 5.76 | 4.79 | .97 |
| Travel time | 5.26 | 6.07 | 4.72 | 1.35 |
| State of vehicle | 5.10 | 5.84 | 4.93 | .91 |
| Congestion | 5.02 | 5.64 | 4.43 | 1.21 |
| Unfamiliarity | 4.56 | 6.03 | 4.88 | 1.15 |
| Diversions | 4.45 | 5.88 | 4.78 | 1.10 |
| Self-esteem | 4.25 | 5.78 | 4.20 | 1.58 |

Note.—For purposes of comparison, only those items that had high factor loadings and were common to both trip purposes were used to calculate the factor scores.
^a Listed in order of importance.
^b Highest possible score = 7.00.

edly more satisfactory than public transport for protection while waiting, as shown in Tables 1 and 2, but does not have much advantage for the vehicle unaffected by weather item. These differences are noteworthy because they pertain to the so-called trade-off hypothesis that suggests passenger exposure to bad weather while waiting for service may be offset by the imperviousness of public transport to such weather. Apparently, the "go in any weather" advantage sometimes attributed to public transport is not a strongly held opinion by the respondents. Thus, it may not effectively counterbalance the definite dissatisfaction with public transport regarding the protection-from-weather-while-waiting item.

Cost

The cost factor, including the four items concerning transport cost, is rated as fourth in importance for both the work and the non-work trip. The respondents saw a distinct advantage—low cost—in the automobile versus public transport mode for both the work and nonwork trips.

For many consumers, the following may be hypothesized: (a) they have made a decision to have a car (based on several considerations); (b) thus, only the variable transportation costs are significant for modal choice decisions; and (c) they view the car to be equal to or less costly than public transit considering variable costs. The latter hypothesis is supported by evidence from other studies, such as Survey Research Center data that showed that of individuals who felt they had a choice between auto and public modes 94 believed the car to be equal to or less costly than the common carrier, while 75 thought the auto to be more expensive.

This does not mean that cost should be underestimated as one of the determinants of consumer decision making. Smerk has reported studies where significant increases in traffic were found to result from fare reductions (Smerk, 1964). As indicated in the present study, cost is ranked fairly high in importance, and respondents are more satisfied with the cost of auto than with public transit. A somewhat analogous situation exists in industry with respect to employee percep-

TABLE 5
PERCENTAGE SATISFIED WITH AUTO FOR "SHORTEST TIME" ITEM COMBINING TRIP PURPOSES

| Category | % satisfied |
|-------------------|-------------|
| Income | |
| \$6,000 | 60 |
| \$6,000-9,999 | 75 |
| \$10,000 and over | 73 |
| Age | |
| Under 35 | 76 |
| 35-55 | 69 |
| Over 55 | 59 |
| Race | |
| White | 72 |
| Other | 60 |
| Dwelling | |
| Single unit | 81 |
| Apartment | 63 |

tions of the importance of wage levels. Such factors as job security and opportunity for advancement have been consistently ranked higher. However, when asked what factors contribute most to dissatisfaction, wages are mentioned frequently. These results may be interpreted as suggesting that the present level of wages for most employees is high enough so that other factors are more crucial. However, if wages are dropped significantly, or if they are perceived to be inequitably determined, it is likely that their importance increases substantially, as suggested by the dissatisfaction evidence. Assuming reasonable service equivalence, perhaps the same phenomenon prevails if there is a significant increase in the cost of auto operation (e.g., fees and tolls) or decrease in the cost of common carrier (e.g., subsidy). It is also suggested that of these two possibilities, increasing the cost of using autos is more effective than decreasing the cost of the common carrier, since neither the present variable cost of driving nor common carrier cost is apparently considered seriously burdensome by many travelers. Of course, other considerations might make it impractical to attempt such a change, for example, political infeasibility.

Table 6 indicates for the low cost item the percentage satisfied with the auto and with public transit categorized by demographic variables. This table shows quite clearly the much smaller percentage satisfied with the

TABLE 6

PERCENTAGE SATISFIED WITH COST OF TRIPS

| Category | Auto | Public transit | Difference |
|---|------|----------------|------------|
| Income | | | |
| <\$6,000 | 54% | 26% | 28% |
| \$6,000-9,999 | 65% | 37% | 28% |
| \$10,000 and over | 67% | 38% | 29% |
| Education | | | |
| <High school | 63% | 31% | 32% |
| High school | 73% | 29% | 44% |
| >High school | 69% | 46% | 23% |
| Distance from central business district | | | |
| <3 miles | 58% | 22% | 36% |
| 3-5 miles | 70% | 43% | 34% |
| >5 miles | 83% | 45% | 38% |

low cost of public transit as well as the pattern of those with lower income, lower education, and residing closer to the central business district showing relatively less satisfaction. Presumably, the less affluent group would be most affected by lowering the costs of public transit or raising the cost of auto travel. However, this group already has a relatively high use rate of public transit and relatively low rate for the auto compared to the more wealthy. Thus, cost changes might help in changing use patterns, but rather less well than might be desired. Cost or changes in cost may be included, however, as one of the elements in the transport package (including travel time, convenience, and weather protection) that determine auto favorability over public transit.

State of Vehicle

Tables 1 and 2 indicate that the items in this factor "clean vehicle" and "new modern vehicle" have moderate importance, with the former slightly more important.

The automobile again emerges as more satisfactory to respondents than does public transport for the newness and the cleanliness of the vehicle. The average difference (.91 and .85) is not as large as it is for most other factors, suggesting possibly that changes in the state of the vehicle would have a relatively small effect on consumer choice.

Convenience

Two items that ranked quite high in importance defined a "convenience" factor. They were "avoid changing vehicle" (6.10 and 5.99) and "avoid waiting" (5.86 and 5.40). In both cases, the auto was seen as satisfactory by a substantially greater percentage of respondents than was public transit as indicated in Tables 1 and 2.

Satisfaction as measured by the combined trip purpose percentage increased with income and distance of residence from the central business district, as illustrated by the "avoid changing vehicle" item in Table 7. It is also evident that differences in satisfaction between auto and public modes are consistently large (31-37%) in favor of the auto for all demographic categories. Thus, it appears there is a tendency for higher income suburbanites to be more satisfied with either mode of travel than the low-income close to central business district group, perhaps simply reflecting a "more satisfied" general syndrome. Both groups obviously agree, however, that the auto is significantly more satisfactory than public modes. Improvement in convenience seems a high priority factor in possible attempts to divert use patterns of the two groups considering both importance and satisfaction data.

Unfamiliarity

Similar to the state of vehicle factor the items in this factor "avoid riding with

TABLE 7

AVOID CHANGING VEHICLE SATISFACTION
(PERCENTAGE SATISFIED)

| Category | Auto | Public transit | Difference |
|---|------|----------------|------------|
| Income | | | |
| <\$6,000 | 74% | 43% | 31% |
| \$6,000-9,999 | 88% | 57% | 31% |
| \$10,000 and over | 88% | 51% | 37% |
| Distance from central business district | | | |
| <3 miles | 67% | 31% | 36% |
| 3-5 miles | 84% | 47% | 37% |
| >5 miles | 92% | 59% | 33% |

TABLE 8

AVOID UNFAMILIAR AREA SATISFICATION
(Percentage Satisfied)

| Category | Auto | Public transit | Difference |
|---|------|----------------|------------|
| Income | | | |
| <\$6,000 | 57% | 37% | 20% |
| \$6,000-9,999 | 76% | 60% | 10% |
| \$10,000 and over | 82% | 51% | 31% |
| Education | | | |
| <High school | 62% | 44% | 18% |
| High school | 77% | 52% | 25% |
| >High school | 78% | 57% | 21% |
| Distance from central business district | | | |
| <3 miles | 62% | 44% | 18% |
| 3-5 miles | 77% | 52% | 25% |
| >5 miles | 78% | 57% | 21% |

strangers" and "avoid unfamiliar area" have some to moderate importance (4.07 and 5.17).

The auto allows a person to travel with whom he pleases and to stay in familiar territory. Public transportation usually does not provide this sort of choice. As might be expected, the auto was rated significantly better in satisfying this requirement as indicated in Tables 1 and 2. Moreover, satisfaction with the unfamiliarity factor varies with income, education, and distance from the central business district. For example, Table 8 shows those who are less highly educated, have lower incomes, and live closer to the central business district tend to be less satisfied in being able to avoid unfamiliar areas in their recent transportation. They perhaps are more likely to begin or end their trips in unfamiliar areas. In traveling to work or to shop, they may tend to go outside of their own areas, perhaps reflecting the perception that jobs and/or prices are more attractive elsewhere. A person traveling in such a vicinity is likely not to be fully familiar with each neighborhood. Again, the auto apparently provides a more satisfactory means of avoiding or lessening this problem.

Self-Esteem

The self-esteem factor was defined by several items including feeling independent, drive vehicle yourself, pride in vehicle, and satis-

faction of owning vehicle. This discussion is limited to the most important item, feeling independent. This item was rated as very important (5.52) on the average, and about 65% of the respondents felt it to be very important or of greatest importance for the work trip (55% for nonwork trip).

Table 9 shows the percent of respondents satisfied with auto and with public transit as it varies with income, education, and distance from the central business district. The largest differences occur for those with higher incomes, more education, and who live farther from the central business district, which is interpreted as additional support for the contention that it will be very difficult to provide an alternative form of public transportation as satisfactory as the auto for the more affluent group.

Congestion and Diversions

The findings of the Baltimore and Philadelphia study support the results of the Survey Research Center which indicate that congestion (travel when traffic is light, uncrowded vehicle) and diversions (take along family and friends, look at scenery, listen to radio, ride with people who chat, friendly people) are relatively unimportant in choosing between auto and common carrier.

Public transport is seen as less satisfactory than the auto for providing the user with op-

TABLE 9

FEELING INDEPENDENT SATISFICATION
(Percentage Satisfied)

| Category | Auto | Public transit | Difference |
|---|------|----------------|------------|
| Income | | | |
| <\$6,000 | 57% | 30% | 27% |
| \$6,000-9,999 | 72% | 38% | 34% |
| \$10,000 and over | 82% | 35% | 47% |
| Education | | | |
| <High school | 61% | 38% | 23% |
| High school | 74% | 32% | 42% |
| >High school | 78% | 30% | 48% |
| Distance from central business district | | | |
| <3 miles | 49% | 25% | 24% |
| 3-5 miles | 71% | 45% | 26% |
| >5 miles | 81% | 35% | 46% |

portunities for diversion and to avoid congestion. For example, for the work trip the auto received an average satisfaction level of 6.27 (high satisfaction) for the uncrowded vehicle item, whereas the public transportation alternatives were rated as only providing moderate satisfaction (4.38 average). However, this large difference in satisfaction is counterbalanced by the relatively low importance of this factor when considering possible implications for changing travel behavior.

DISCUSSION

In assessing ways of modifying consumer decision making and travel behavior, it is important to measure resistance to change levels for various groups. One method is to examine the differences in average satisfaction levels between public transit and auto considering various attributes and their importance. Another approach is to examine percentage differences between various demographically defined subgroups' perceived satisfaction with public transit. The larger differences (along with relatively high importance) for characteristics would probably indicate a high level of resistance to change for the demographic group considered.

Such important characteristics with large satisfaction differences for the Baltimore and Philadelphia respondents were many. Perception of travel time, susceptibility to weather, avoidance of changing vehicles, avoidance of waiting, avoiding the unfamiliar, and perceiving an independent feeling were more significant than cost, reliability of destination achievement, state of vehicle, congestion, and diversions in choosing between auto and public transit according to the evidence presented.

The most satisfied group of transport consumers and perhaps the group most difficult to change were the middle and upper income suburbanites living more than 3 miles from the central business district. They are, of course, the ones who possess one or more autos and find it more feasible and satisfying to use them for work and nonwork trips. These same respondents, though they tended to be fairly well satisfied with public transit for many characteristics, indicate strong pref-

erences for the auto when considering such factors as feeling independent, saving time, avoiding change of vehicles, and avoiding unfamiliar areas.

The affluent group presumably would be a key target of persuasion campaigns to use public transit, since they use it the least. The data presented herein should be useful in identifying the content and strategy of such campaigns. Although evidence in this study indicates a consistent preference for auto over public transit, it is clear that there are large differences in the magnitude of this preference when various mode attributes are considered. If these magnitude differences are coupled with significant differences in the importance of attributes, implications for the direction and likely success of future attempts to change mode use patterns become clearer. The data in this study suggest that it will take substantial ingenuity and resources to make public transit an attractive alternative for those predominantly using the automobile for their work and nonwork trips.

REFERENCES

- ACKOFF, R. L. *Individual preferences for various means of transportation*, Philadelphia: University of Pennsylvania, 1965.
- GILAT, H. I. *Investigation of present methods for establishing a modal split to travel estimates*. Baltimore, Md.: Parsons, Brinckerhoff, Quade and Douglas, September-October 1963.
- LANSING, J. B., MUELLER, E., & BARTH, N. *Residential location and urban mobility*. Ann Arbor, Michigan: Survey Research Center, Institute for Social Research, University of Michigan, June 1964. (Prepared for United States Bureau of Public Roads.)
- MAHONEY, J. F. *A survey to determine factors which influence the public's choice of mode of transportation*. Boston: Joseph Napolitan Associates, 1964.
- MINER, J. B. *Personnel and industrial relations*. New York: Macmillan, 1969.
- SMERK, G. M. Three experiments in urban transportation. *Business Horizons*, 1964, Summer, 39-47.
- STANFORD RESEARCH INSTITUTE. *The value of time for passenger cars: Further theory and small-scale behavioral studies*. (Draft copy for client review; prepared for United States Bureau of Public Roads, SRI Project IM-4393, Contract CPR-11-8028, 1965.) Stanford, Calif.: Author, 1965.
- STOCKFORD, L., & BISSELL, H. W. Establishing a graphic rating scale. *Personnel*, 1949, 26, 94-116.

(Received December 6, 1968)

CORRELATES OF EMPLOYEE EVALUATIONS OF PAY INCREASES

J. R. HINRICHS¹

IBM, White Plains, New York

Within a large, white-collar, industrial population, average perceptions of small, average, or large increases in salary formed a relatively constant function of level of current salary. The analogy to the psychophysical Weber/Fechner model, while explaining much of the variance in perceptions of salary increases was not complete. Additional variability was related to a series of demographic variables, with higher dollar expectations registered by college-educated employees versus noncollege, younger employees versus older, exempt employees versus nonexempts, and among nonexempts by males versus females. The results suggest that probable earnings potential, in addition to current earnings level, contributes variance to differences in perceptions of equitable salary increases.

A recent review of research in the area of compensation (Opsahl & Dunnette, 1966) illustrates that very little is really known about the incentive value of money. Most of the data that have been published are exclusively psychological, usually dealing primarily with various aspects of satisfaction with pay, pay expectations, or rated importance of money in comparison with other job elements (e.g., the studies reviewed in Herzberg, Mausner, Peterson, & Caldwell, 1957; and Vroom, 1964). While job attitudes with respect to pay may be important correlates of behavior, without some basis for relating attitudes to objective pay in terms of compensation dollars there is only limited direct utility from these studies for the more efficient administration of compensation. And, on the other hand, what "hard" data there are dealing with compensation in turn are often not tied to perceptions (e.g., Haire, Ghiselli, & Gordon, 1967).

Studies that do attempt to evaluate the bridge between actual compensation and attitudes tend to deal with perceptual data in the form of general satisfaction and overall expectations regarding pay, evaluating these perceptions for individuals at differing compensation levels (e.g., Andrews & Henry, 1963; or Lawler & Porter, 1963). The most prevalent interpretation emerging from these analyses is that satisfactions regarding pay

are less a function of absolute current levels of pay than they are of motivational variables such as goals and expectations, or of background factors such as sex, age, or education which serve as reference points for building goals and expectations.

Most frequently, psychological studies of compensation deal only tangentially, if at all, with the extent to which established salary administration policies and practices serve as normative factors shaping perceptions of pay. Past experience with pay, one would expect, is probably a crucial variable in shaping future expectations; in large measure one would expect that demographic variables correlate with perceptions of pay primarily as a result of common variance attributable to differences in actual pay treatment. Thus, it would seem desirable for research in the area of compensation to take a more critical and comprehensive look at the motivational and behavioral correlates of actual dollar amounts of pay, under the assumption that these data will capture the essence of the compensation system that shapes the individual's psychological world in this critical area.

An organization's pay practices and policies are expressed most demonstrably as changes in pay. The key variable in understanding perceptions of pay, if we wish to deal with the norms of the pay system as it shapes these perceptions, must be incremental pay, or raises. And the research paradigm of classical psychophysics dealing with perceptions

¹Requests for reprints should be sent to J. R. Hinrichs, IBM Corporation, 112 East Post Road, White Plains, New York 10601.

of incremental stimulus intensity would seem completely appropriate to the study of perceptions of pay.

A recent study of Zedeck and Smith (1968) has used the psychophysical method of limits to evaluate perceptions of equitable salary among junior executives and secretaries within a Midwestern academic institution. While the study suggests the utility of this method in providing an understanding of perceptions of pay equity, due to limitations in sample size, methodological problems in using the survey technique for data collection, and the restricted range of occupations and backgrounds of Ss in the study, it is impossible to draw any conclusions regarding perceptions of equitable salary increments among industrial employees.

The present study is an attempt to provide data that bridge both the area of perceptions regarding pay and hard data on actual levels of compensation for a large and diverse industrial population. The study focuses directly on attitudes regarding pay increases and deals with two overlapping propositions: (a) that on the average, expectations regarding compensation follow a relatively consistent pattern that is largely a function of an individual's current absolute level of earnings; and (b) beyond this general patterning, pay expectations are further influenced by normative factors that are largely a function of an individual's personal situation and background.

The first proposition assumes that some form of Weber-Fechner relationship exists with regard to money; that is, that perceptions of what would be a just noticeable increment in earnings is largely a constant function of current earnings. Or, in more prosaic terms, it assumes that the more money an individual makes, the more of an increment in salary it will take to stir his feelings, and that this increment is a relatively constant fraction of his current level of earnings.

The second proposition says that while, on the average, just noticeable pay increments will reflect some relatively constant function of current earnings, there will be considerable variability around this average and that this variability will be associated with a number of factors describing the individual's back-

ground and particular situation. Based upon the analyses by Andrews and Henry (1963), Klein and Maher (1966), and by Penzer (1969), one might expect level of education to be one such factor influencing an individual's expectation. The analysis by Lawler and Porter (1963) across various levels of management would suggest that organizational level would be another such variable. Similarly, we might expect age, sex, and occupation to be related significantly to earnings expectations.

METHOD

The data in this study come from an employee attitude survey administered to roughly 1,500 employees in a large and geographically dispersed industrial organization. Survey participants were involved in a diversity of white-collar occupations. Three-quarters of them were male and one-fourth female. Forty-four percent were nonexempt employees, 38% nonmanagerial exempt employees, and 18% were managers. All were paid on a salary basis. While the questionnaire consisted of over 200 items dealing with a broad spectrum of attitude and opinion areas, only a few items were used in the analyses reported here.

The key item in the analysis consisted of a listing of hypothetical dollars-per-month increases ranging from \$1 per month to \$1,000. Instructions stated:

Everyone will agree that a monthly increase of \$1,000 would be an "extremely large" salary increase. At the same time, an increase of \$1 per month would most likely be viewed as a "just barely noticeable increase." Somewhere between these extremes people would view different dollar amounts as representing "large" or "average" or "small" increases in salary.

To provide data dealing with a theory of "rewards," we would like you to think about how you would view different salary increases. Please think realistically. We would like you to divide the following list of dollar amounts into five segments representing dollar increases in monthly salary which you would tend to view as falling in each of these categories:

1. Extremely large increases in monthly salary; I would be flabbergasted.
2. Large salary increases; I would be pleasantly surprised.
3. Neither large nor small increases; about average.
4. Small salary increases; I would be somewhat disappointed.
5. Just barely noticeable salary increases; ones which essentially would not be viewed as an increase at all.

The salary increments included in the list were \$1, \$5, \$10, \$15, \$20, \$25, \$30, \$40, \$50, \$60, \$80, \$100,

\$125, \$150, \$175, \$200, \$250, \$300, \$400, \$500, \$1,000. The data that were recorded were the midpoints of the four intervals that each respondent indicated divided this series into the five different categories of dollar increases for him.

Another key item in the analysis asked: "What is your current monthly salary?"

1. Under \$400

2. \$400 to \$499

3. \$500 to \$599

4. \$600 to \$699

5. \$700 to \$799
6. \$800 to \$899

7. \$900 to \$1099

8. \$1100 to \$1399

9. \$1400 to \$2000

10. Over \$2000

Various demographic variables such as sex, age, education, and organizational level were also used in the analysis.

RESULTS

For each individual, the midpoints of the salary range that he indicated as separating "just barely noticeable" from "small" salary increases, "small" from "average," "average" from "large," and "large" from "extremely large" increases were recorded. Table 1 shows the means and standard deviations from these midpoints for male employees at each of the 10 different levels of current monthly earnings. Table 2 presents comparable data for females. The "percentage of current salary" presented for each of these dividing points is merely the mean of the hypothetical increases indicated by people at each of the current salary levels over the midpoint of the class

interval for that salary level; for the top and bottom categories in the current monthly salary distribution, which were open-ended categories in the questionnaire, the midpoint of the interval was assumed for purposes of computing the percentages.

As these tables suggest, the average dollars that were indicated as the dividing point between "average" and "small" salary increases and between "small" and "just barely noticeable" increases, etc., generally increase in a monotonic fashion with total salary. The "Weber fraction" is relatively stable, especially for the women, regarding the salary that would be perceived as "small" versus "just barely noticeable." There is, however, a slight though consistent tendency for these ratios of salary increase to salary level to decrease as level of earnings increases, for all of the categories of perceived increases. This suggests that perhaps the pure psychophysical model may not hold but that perhaps other factors are also systematically affecting these perceptions. The most stability to the psychophysical model occurs in the range between \$400 and \$2,000 current salary, and the data at either extreme of the distribution should probably be deemphasized because the class interval was not completely specified in the questionnaire, and the midpoint salary was

TABLE 1
PERCEPTIONS OF MINIMAL AND AVERAGE SALARY INCREASES AS A
FUNCTION OF CURRENT EARNINGS FOR MALES

| Current monthly salary | N | "Just barely notice- able" (versus "small") increases | | | "Small" (versus "average") increases | | | "Average" (versus "large") increases | | | "Large" (versus "very large") increases | | |
|------------------------------|-----|---|----------|---------------------------|---|----------|---------------------------|---|----------|---------------------------|--|----------|---------------------------|
| | | \bar{X} | σ | % of current salary | \bar{X} | σ | % of current salary | \bar{X} | σ | % of current salary | \bar{X} | σ | % of current salary |
| Under \$400 ^a | 18 | \$18 | 6 | 5.1 | \$32 | 10 | 9.1 | \$76 | 22 | 22 | \$159 | 59 | 45 |
| \$400-499 | 96 | \$19 | 7 | 4.2 | \$35 | 13 | 7.8 | \$67 | 26 | 15 | \$152 | 65 | 34 |
| \$500-599 | 107 | \$20 | 8 | 3.7 | \$38 | 17 | 6.9 | \$78 | 39 | 14 | \$179 | 118 | 32 |
| \$600-699 | 95 | \$21 | 10 | 3.3 | \$39 | 14 | 6.0 | \$77 | 32 | 12 | \$171 | 88 | 26 |
| \$700-799 | 113 | \$25 | 11 | 3.3 | \$47 | 18 | 6.3 | \$95 | 33 | 13 | \$207 | 101 | 28 |
| \$800-899 | 122 | \$27 | 11 | 3.2 | \$51 | 17 | 6.0 | \$100 | 35 | 12 | \$196 | 80 | 23 |
| \$900-1099 | 193 | \$31 | 14 | 3.1 | \$59 | 21 | 5.9 | \$115 | 38 | 12 | \$222 | 84 | 22 |
| \$1100-1399 | 160 | \$38 | 16 | 3.0 | \$71 | 22 | 5.7 | \$135 | 40 | 11 | \$261 | 108 | 21 |
| \$1400-2000 | 173 | \$50 | 22 | 2.9 | \$91 | 29 | 5.4 | \$165 | 43 | 10 | \$310 | 111 | 18 |
| Over \$2000 ^b | 34 | \$56 | 24 | 2.2 | \$107 | 35 | 4.3 | \$191 | 50 | 8 | \$364 | 111 | 15 |

^a Midpoint assumed at \$350.
^b Midpoint assumed at \$2500.

TABLE 2
PERCEPTIONS OF MINIMAL AND AVERAGE SALARY INCREASES AS A
FUNCTION OF CURRENT EARNINGS FOR FEMALES

| Current monthly salary | N | "Just barely notice- able" (versus "small") increases | | | "Small" (versus "average") increases | | | "Average" (versus "large") increases | | | "Large" (versus "very large") increases | | |
|------------------------------|-----|---|----------|---------------------------|---|----------|---------------------------|---|----------|---------------------------|--|----------|---------------------------|
| | | \bar{X} | σ | % of current salary | \bar{X} | σ | % of current salary | \bar{X} | σ | % of current salary | \bar{X} | σ | % of current salary |
| Under \$400 ^a | 30 | \$14 | 5 | 3.9 | \$28 | 13 | 8.0 | \$61 | 28 | 17 | \$159 | 126 | 45 |
| \$400-499 | 113 | \$16 | 6 | 3.5 | \$29 | 8 | 6.4 | \$60 | 20 | 13 | \$137 | 57 | 30 |
| \$500-599 | 81 | \$19 | 7 | 3.5 | \$35 | 12 | 6.4 | \$68 | 27 | 12 | \$145 | 71 | 26 |
| \$600-699 | 70 | \$22 | 8 | 3.4 | \$40 | 13 | 6.2 | \$79 | 31 | 12 | \$168 | 100 | 26 |
| \$700-799 | 35 | \$26 | 10 | 3.5 | \$49 | 16 | 6.5 | \$95 | 27 | 13 | \$178 | 54 | 24 |
| \$800-899 | 14 | \$26 | 10 | 3.1 | \$48 | 15 | 5.6 | \$90 | 21 | 11 | \$161 | 38 | 19 |
| \$900-1099 | 12 | \$35 | 11 | 3.5 | \$60 | 15 | 6.0 | \$104 | 28 | 10 | \$261 | 230 | 26 |
| \$1100-1399 | 9 | \$47 | 25 | 3.7 | \$82 | 27 | 6.5 | \$145 | 31 | 12 | \$261 | 49 | 21 |
| \$1400-2000 | 3 | — | — | — | — | — | — | — | — | — | — | — | — |
| Over \$2000 ^b | 0 | — | — | — | — | — | — | — | — | — | — | — | — |

^a Midpoint assumed at \$350.
^b Midpoint assumed at \$2500.

merely assumed in computing the fraction. Omitting these two extreme categories, the percentage increase cited as dividing "just barely noticeable" from "small" ranges between 4.2% and 2.9% for males, with an average of 3.3%; the range is between 3.1% and 3.7% for females with an overall average of 3.5%. The overall average "small" versus "average" increase split is 6.1% for males and 6.3% for females. It is interesting to compare this with the average K of 6.2% for secretaries reported by Zedeck and Smith (1968, p. 345).

While there appears to be a fair amount of stability in defining these percentages for small or for average increases, it is clear that there is a less stable perception of what represents a large or a very large increase. One might suspect that the norms of the organization with regard to salary administration policy and treatment engender certain fairly stable expectations of the average, but that evaluations of what constitutes a large increase are built largely on wishful thinking, conjecture, and a considerable element of random response.

Figures 1 and 2 further clarify these relationships. These figures are log-log plots of current monthly salary versus perceptions of mean increases, separately for males and for

females. The trend lines for perceived small and average increases are quite linear, conforming to the logarithmic model as suggested by Haire et al. (1967), while the trends for large and very large increases are less regular. Based upon these plots and the data in the tables, it would seem that average perceptions of various salary increases in large measure do follow a lawful relationship to current levels of salary, and that these perceptions conform to a considerable degree to the Weber-Fechner type of relationship. The slight S cast to the curves in Figures 1 and 2 highlights the apparently systematic deviation from the psychophysical model discussed above. As Tables 1 and 2 indicate, there is considerable variability around the average perceptions of increases used in computing these Weber fractions. The data in Tables 3 through 6 illustrate that to a certain extent, various demographic factors are associated with this variance and account for additional variability beyond that which may be attributed to current levels of salary. For each of the current monthly salary levels the distribution of dollar increases that individuals within each level cited as separating small from average increases were divided as closely as possible into equal thirds. Tables 3-6 contrast individuals falling in the upper third of the

distribution of perceived increases within their respective current salary level with those falling in the middle third and those falling in the lower third for various demographic classifications.

Table 3 indicates that when current salary level is controlled, college graduates are significantly more apt to have a higher perception of what is an "average" salary increase than are noncollege graduates, and the trend is such as to suggest that possibly this is more true of college graduates at the bachelor's degree level than of individuals with gradu-

ate degrees. The data in Table 4 reflect differences in the perceptions of average salary increases with age and indicate an increasingly large fraction of respondents with relatively low salary expectations in successively older groups, while younger employees are more apt to report high expectations. Table 5 suggests that employees exempt from the provisions of the Fair Labor Standards Act are perhaps somewhat more apt to have higher salary increase expectations than are non-exempt employees. Based upon the data in Table 6, there are no evident differences be-

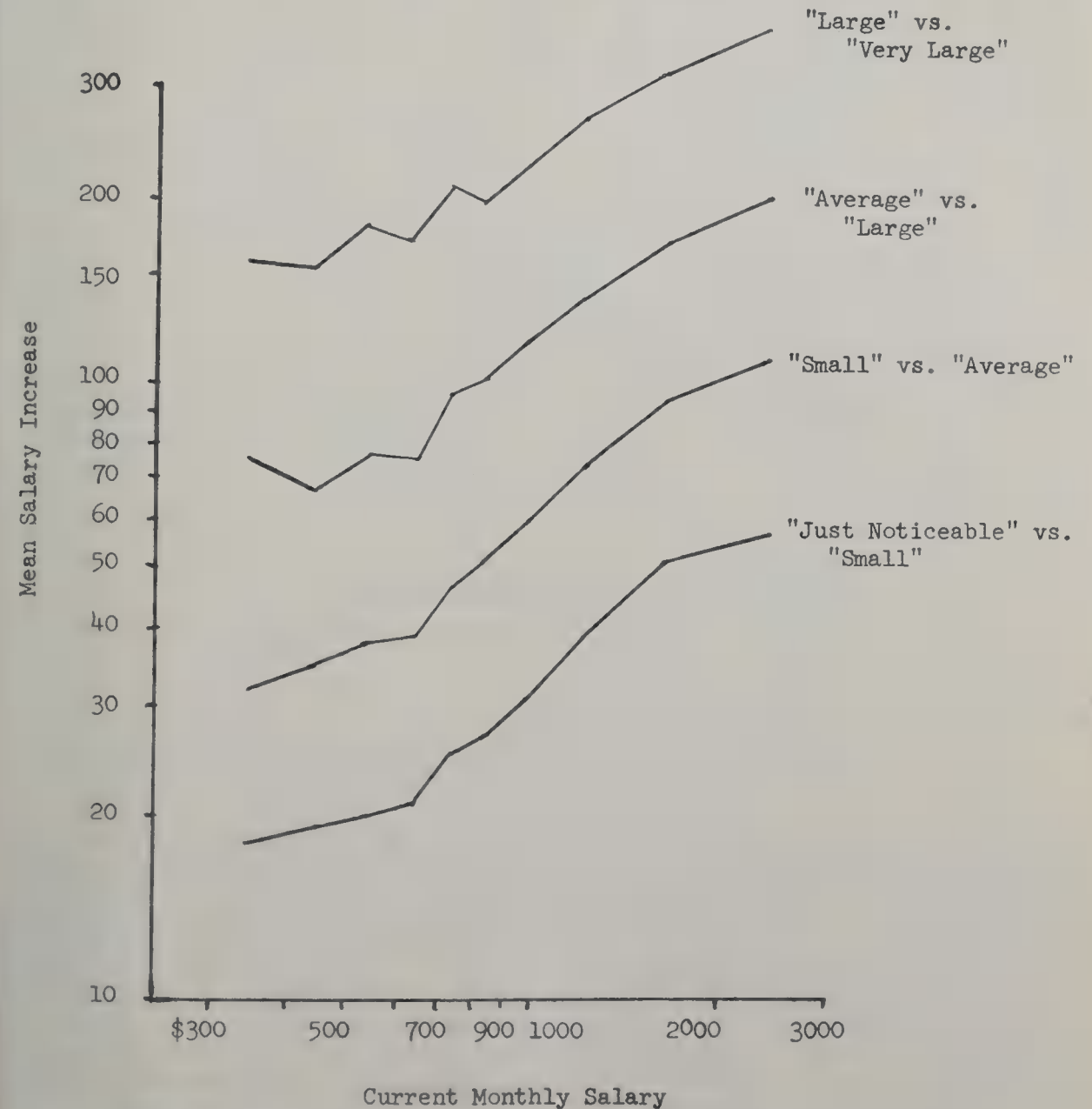


FIG. 1. Perceptions of salary increments as a function of current salary (males only).

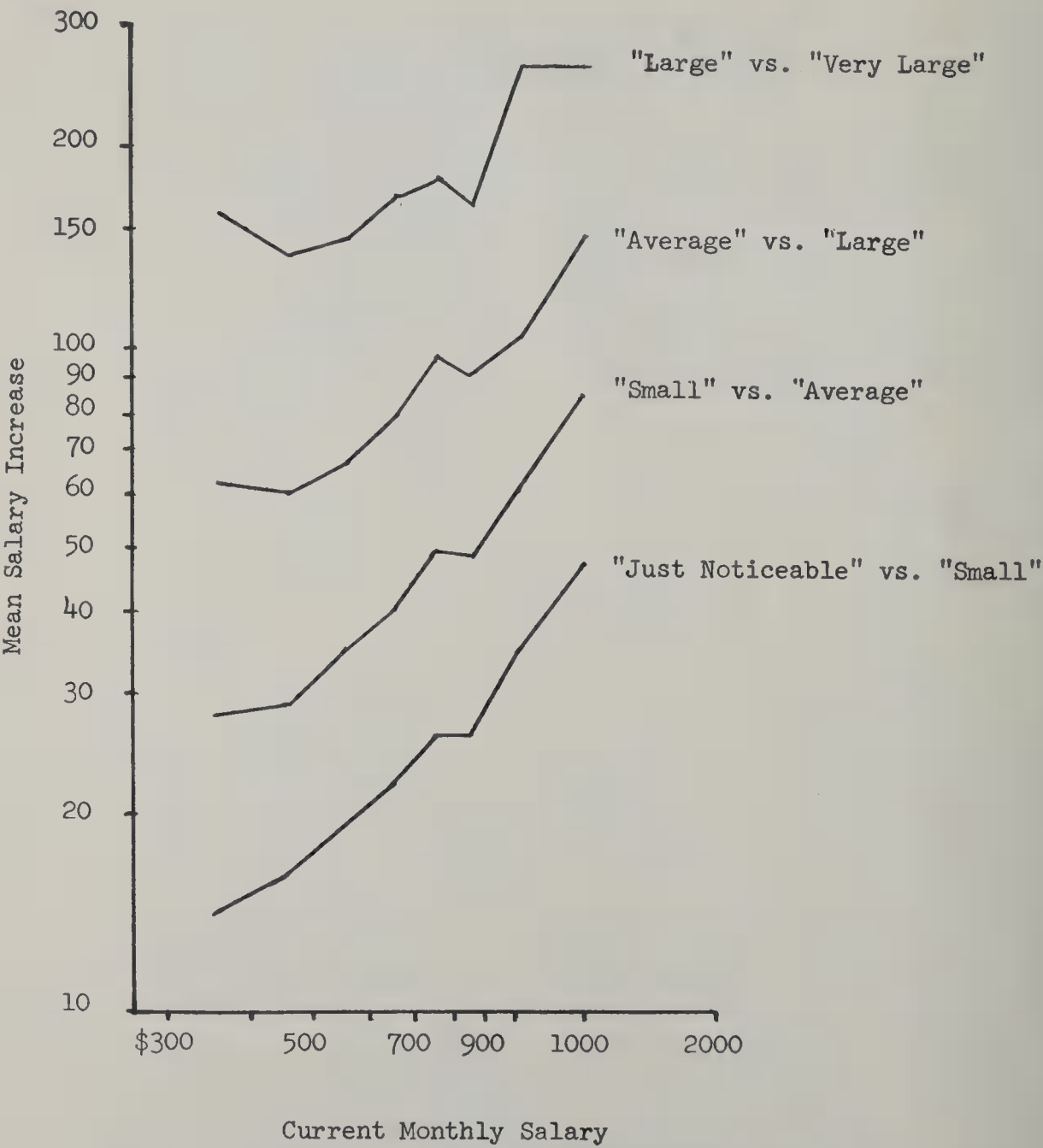


FIG. 2. Perceptions of salary increments as a function of current salary (females only).

tween males and females in relative expectations regarding salary increases.

The types of demographic variables investigated in Tables 3–6 are clearly not mutually exclusive, and a series of two-variable cross-tabulations such as this goes only part way in clarifying which are the most critical factors in explaining perceptions of salary increases or in identifying possible interactions among the variables. However, an analysis using AID—a computer program designed to iden-

tify the optimal combinations of an array of independent variables explaining a selected dependent variable (Sonquist & Morgan, 1964)—does adequately handle this problem. Figure 3 summarizes the results of this analysis and shows the “tree” of relationships that sequentially explains the most variance in the dependent variable of relative perceptions of average salary increases, with current actual salary level controlled.

The most “critical” demographic variable—

TABLE 3

RELATIVE PERCEIVED "AVERAGE" INCREASE
VERSUS EDUCATION

| Relative perception of "average" increase | Educational level | | |
|--|-------------------|-----------------|-------------------------|
| | Non- college | Bach- elor's | Grad- uate degree |
| Upper one-third, within earnings group | 241 (30.9%) | 268 (48.7%) | 52 (41.9%) |
| Middle one-third | 252 (32.3%) | 147 (26.7%) | 39 (31.5%) |
| Lower one-third, within earnings group | 287 (36.8%) | 135 (24.6%) | 33 (26.6%) |
| Total | 780 (100%) | 550 (100%) | 124 (100%) |

Note.—Current earnings held constant. $\chi^2 = 46.79, p < .01$

that is, the one that explains the most variance in the dependent variable—is education. In line with the findings of Klein and Maher (1966) and of Penzer (1969), college-educated employees have significantly higher expectations regarding salary increases than do noncollege graduates.

Age is the next most important variable accounting for differences in perceptions of increases. For both college and noncollege employees, younger employees have significantly

TABLE 4

RELATIVE PERCEIVED "AVERAGE" INCREASE
VERSUS AGE

| Relative perception of "average" increase | Age | | |
|--|----------------|----------------|----------------|
| | Under 30 | 30-39 | 40 or older |
| Upper one-third, within earnings group | 224 (41.2%) | 252 (38.4%) | 76 (29.6%) |
| Middle one-third | 189 (34.7%) | 183 (27.8%) | 70 (27.2%) |
| Lower one-third, within earnings group | 131 (24.1%) | 222 (33.8%) | 111 (43.2%) |
| Total | 544 (100%) | 657 (100%) | 257 (100%) |

Note.—Current earnings held constant. $\chi^2 = 33.39, p < .01$.

TABLE 5

RELATIVE PERCEIVED AVERAGE INCREASE
VERSUS EMPLOYEE STATUS

| Relative perception of average increase | Non- exempt | Exempt non- manager | Manager |
|--|----------------|---------------------------|----------------|
| Upper one-third, within earnings group | 222 (34.9%) | 238 (42.0%) | 102 (38.5%) |
| Middle one-third | 204 (32.1%) | 164 (28.9%) | 74 (27.9%) |
| Lower one-third, within earnings group | 210 (33.0%) | 165 (29.1%) | 89 (33.6%) |
| Total | 636 (100%) | 567 (100%) | 265 (100%) |

Note.—Current earnings held constant. $\chi^2 = 7.27, p < .10$.

higher expectations than do older. The optimal split in explaining perceptions of salary increases (based on data coded in 5-yr. class intervals) occurs at age 25 for the noncollege group, and at age 30 for the college graduates, suggesting that perhaps the key factor is the individual's number of years of exposure to industrial work and salary administration practices, rather than his chronological age per se.

The sex variable further explains relative perceptions of salary increases only for the younger noncollege group, with males having significantly higher expectations than females.

TABLE 6

RELATIVE PERCEIVED "AVERAGE" INCREASE
VERSUS SEX

| Relative perception of "average" increase | Sex | |
|--|----------------|----------------|
| | Male | Female |
| Upper one-third, within earn- ings group | 428 (38.8%) | 135 (36.8%) |
| Middle one-third | 327 (29.7%) | 115 (31.3%) |
| Lower one-third, within earn- ings group | 347 (31.5%) | 117 (31.9%) |
| Total | 1002 (100%) | 367 (100%) |

Note.—Current earnings held constant. $\chi^2 = .713, ns$.

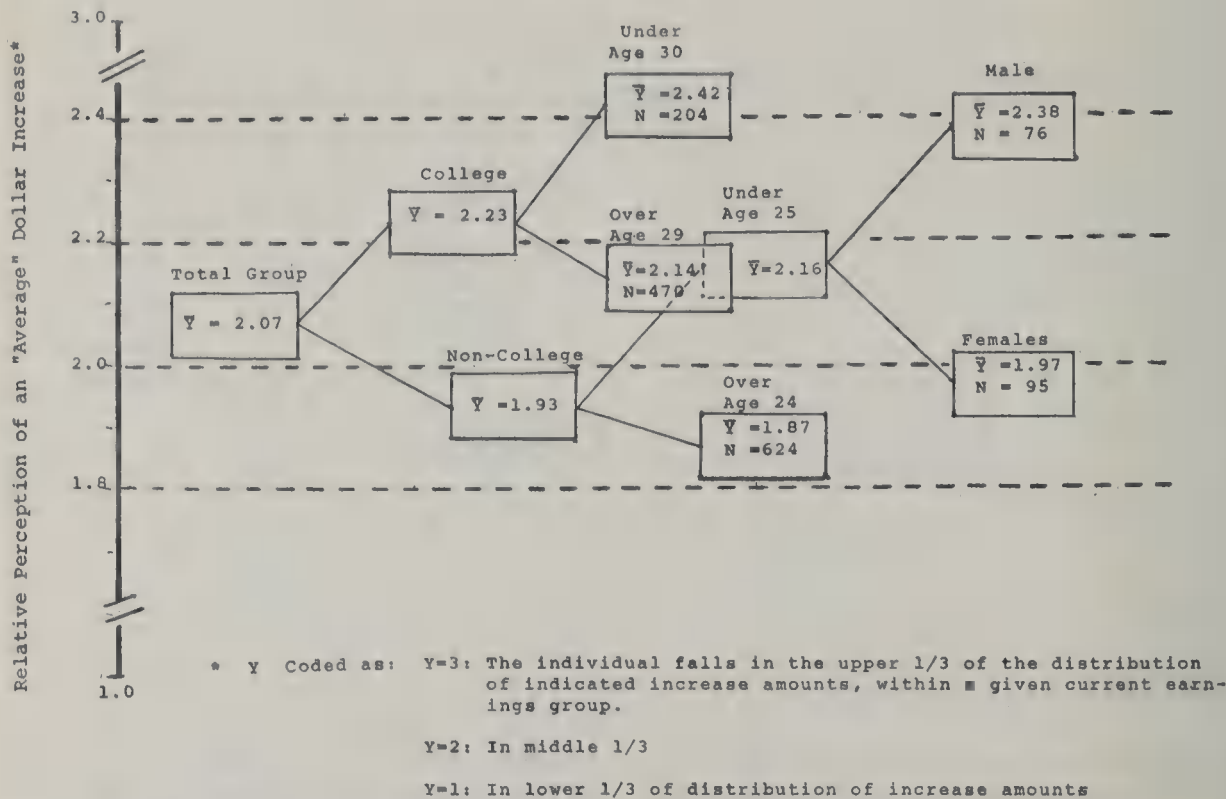


FIG. 3. Demographic characteristics explaining relative perceived "average" salary increases (current earnings held constant).

This finding highlights the importance of assessing interactions in an analysis such as this, as the simple cross-tabulation in Table 6 does not reflect any differences between males and females.

The demographic variable of occupational level—nonexempt versus nonmanagerial versus managers—does not contribute any significant unique explanation of variance in perception of increases, no doubt reflecting the high correlation between this variable and education level and age.

DISCUSSION

In this presentation we have deliberately avoided an analysis of opinion questionnaire data dealing directly with satisfaction with salary and ratings of earnings expectations in preference for our analysis which deals with perceptions regarding actual dollar amounts and demographic characteristics. Hopefully, by maintaining a reasonably concrete referent in the form of actual dollars for the perceptual data that are presented, we have avoided some of the problems of ambiguity of referent

and response biases such as halo and acquiescence that sometimes occur in opinion and satisfaction data.

The picture that emerges from this analysis suggests that an individual's current level of earnings is probably one of the most powerful variables affecting how he perceives a given amount of incremental earnings. Certainly, as common sense would tell us, a fixed salary increase—for example, of \$25 per month—has vastly different meaning to different individuals, largely as a function of their current earnings level. To conclude that differences in satisfaction regarding pay are not related to differences in actual level of pay as do Klein and Maher (1966, p. 205) takes as a "given" the fact of a system of compensation with differing norms regarding money depending upon current level of earnings. (Actually, the current data suggest increasing satisfaction with salary with higher levels of earnings, though these were not presented nor are they necessary to the analysis here.) The current data would suggest that this normative structure—which in actuality can be thought of as represented by the organization's salary struc-

ture—is probably the most powerful variable influencing perceptions regarding money.

Beyond that, however, other situational characteristics, largely demographic in nature, seem to influence employee's perceptions of money within this population. The analysis suggests that college graduates, younger employees, and male employees holding non-exempt positions tend to have relatively high thresholds with regard to the motivational potential of money. That is, one would expect that if current salary levels were held constant, male nonexempt employees would be less satisfied than females, college graduates would be less satisfied than noncollege graduates, younger employees would be somewhat less satisfied than older, and exempt employees overall would tend to be less satisfied than nonexempt.

Within the framework of the psychophysical model regarding the perception of salary increases, we perhaps may view these characteristics as analogous to sources of "constant error" such as occur in laboratory studies.² Such "errors," in addition to operating within any one salary level, could also be expected to demonstrate effects across levels, an effect useful in explaining the apparent tendency for the "Weber fractions" in Tables 1 and 2 to decrease slightly for successively higher salary levels. Since salary levels themselves are correlated with these demographic characteristics, we could reasonably expect somewhat larger "Weber fractions" for relatively low salary levels than for higher levels.

The nature of these trends for demographic variables fits in with an expectancy hypothesis regarding earnings suggesting that those individuals possessing personal characteristics ordinarily associated with relatively high levels of earnings *potential*—college education, youth, males (among nonexempts)—will tend to have higher expectations with regard to earnings than will others, and that these expectations in turn are reflected in a higher

threshold of what is perceived as an acceptable salary increase. On the other hand, people with lower expectations will "settle for less." These trends also fit in with the concept of reference groups, which suggests that satisfactions within industry are dependent to at least some extent on the reference group that the individual uses as a norm for evaluating his own situation. But, as pointed out above, probably even a more potent referent for salary perceptions than an individual's reference group is the very basic compensation issue of "where I am now." By taking both factors into account, it may be possible to build relatively comprehensive models of individuals' perceptions regarding equitable compensation.

REFERENCES

- ANDREWS, I. R., & HENRY, M. M. Management attitudes toward pay. *Industrial Relations*, 1963, 3, 29-39.
- HAIRE, M., GHISELLI, E. E., & GORDON, M. E. A psychological study of pay. *Journal of Applied Psychology Monograph*, 1967, 51(4, Whole No. 636.
- HERZBERG, F., MAUSNER, B., PETERSON, R. O., & CAPWELL, D. F. *Job attitudes: Review of research and opinion*. Pittsburgh: Psychological Service of Pittsburgh, 1957.
- KLEIN, S. M., & MAHER, J. R. Education level and satisfaction with pay. *Personnel Psychology*, 1966, 19, 195-208.
- LAWLER, E. E., III, & PORTER, L. W. Perceptions regarding management compensation. *Industrial Relations*, 1963, 3, 41-49.
- OPSAHL, R. L., & DUNNETTE, M. The role of financial compensation in industrial motivation. *Psychological Bulletin*, 1966, 66, 94-118.
- PENZER, W. N. Educational level and satisfaction with pay: An attempted replication. *Personnel Psychology*, 1969, 22, 185-199.
- SONQUIST, J. A., & MORGAN, J. N. *The detection of interaction effects: A report on a computer program for the selection of optimal combinations of explanatory variables*. Ann Arbor: Survey Research Center, University of Michigan, 1964.
- VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.
- ZEDECK, S., & SMITH, P. C. Psychophysics and equitable payments. *Journal of Applied Psychology*, 1968, 52, 343-347.

(Received December 11, 1968)

² The author is indebted to Allen I. Kraut for his contribution in pointing out his analogy.

EFFECTS OF PERFORMANCE ON LEADERSHIP, COHESIVENESS, INFLUENCE, SATISFACTION, AND SUBSEQUENT PERFORMANCE¹

GEORGE F. FARRIS AND FRANCIS G. LIM, JR.²

Sloan School of Management, Massachusetts Institute of Technology

Two hundred management students role played the "Change of Work Procedure" case in a study designed to determine ways in which performance affects leader behavior. Through changes in the foreman's roles, groups were assigned to a high performance, low performance, or control condition. High past performance was found to increase leader supportiveness, interaction facilitation, goal emphasis, and work facilitation behaviors, as well as member influence, group cohesiveness, and satisfaction. Thus, theories of leadership should consider performance as a cause as well as an effect of leader behavior.

Behavioral scientists (see, e.g., Blake & Mouton, 1964; Likert, 1961; or McGregor, 1960) have argued strongly that leadership behavior affects the performance of subordinates. Evidence for this argument has come from a number of correctional studies, for example, the work at the Institute for Social Research at the University of Michigan in the early 1950s (e.g., Katz & Kahn, 1952, 1960; Katz, Maccoby, Gurin, & Floor, 1951; Katz, Maccoby, & Morse, 1950; or Likert, 1961), and from a few experiments. In one experiment, Jackson (1953) found that when supervisors of work groups were transferred to other groups, the new subordinates perceived them in substantially the same manner as the original group. Apparently the supervisors maintained their style of leadership regardless of characteristics of the group being supervised. In another study, Day and Hamblin (1964) reported that feelings of aggression and the productivity of undergraduate women in a laboratory group varied according to two dimensions of leadership: close versus general and punitive versus non-punitive.

¹ This research was supported in part by a grant from the Alfred P. Sloan Research Fund.

² The authors are grateful for the assistance of Eldon E. Senner and James R. Stinger in various phases of this research. A portion of this research is based upon a dissertation submitted in partial fulfillment of the requirements for the Master of Science degree by the junior author in June 1968.

Requests for reprints should be sent to George F. Farris, Massachusetts Institute of Technology, Sloan School of Management, Room 52-590, 50 Memorial Drive, Cambridge, Massachusetts 02139.

Although the findings of these experiments indicate that leadership can affect performance, the possibility remains that the performance of the subordinates can also affect leadership. The findings of the correlational studies of leadership can be interpreted in this way. Moreover, in a recent longitudinal study, Farris (1969) found consistently stronger relationships between performance and several aspects of "leadership climate" when performance was measured first. "Leadership climate" appeared to follow performance more than performance followed "leadership climate."

A full understanding of leadership behavior requires that it be studied as a dependent as well as an independent variable. To the extent that performance affects leadership, causal interpretations of correlations between leadership and performance should allow for the possibility that leadership behavior is affected by performance. The present study examines experimentally the effects of performance upon four aspects of leadership behavior suggested by Bowers and Seashore (1966): support, interaction facilitation, goal emphasis, and work facilitation. It was predicted that each of these four leadership factors, which have been found to be positively *correlated* with different measures of performance, would be *caused* by performance.

A second set of predictions was concerned with feelings about the group and its discussion process. It was predicted that when the leader was told that his group was "high performing," the leader and subordinates

would feel more satisfied with their group and its discussion process, more cohesive, and able to be more productive in the future, and the subordinates would feel better able to influence the discussion process.

PREDICTIONS

Hypothesis 1: Leaders told that they have high-producing groups will be seen by their subordinates as showing more "good leadership" behavior than leaders told that they have low-producing groups. This fundamental hypothesis of the present study is based upon the assumption that positive correlations found between performance and leadership in past studies are due in part to performance affecting leadership behavior. Performance is predicted to affect leadership in four areas:

Hypothesis 1a: Support. When compared to leaders who are told they have low-producing groups, leaders told they have high-producing groups will be seen by their subordinates as more sensitive to subordinates' needs and feelings, more apt to give recognition for good work, more trustful of the subordinates, less punitive and critical, and less apt to exert unreasonable pressure.

Hypothesis 1b: Goal emphasis. When compared to leaders who are told they have low-producing groups, leaders told they have high-producing groups will be seen by their subordinates as more apt to let subordinates know what is expected from them, maintain high performance standards, stress group pride, and stress being ahead of the competition.

Hypothesis 1c: Work facilitation. When compared to leaders who are told they have low-producing groups, leaders told they have high-producing groups will be seen by their subordinates as more apt to explain suggested job changes and to allow freedom in the work but less apt to decide in detail what shall be done and to impose their own preferred solutions in problem solving.

Hypothesis 1d: Interaction facilitation. When compared to leaders who are told they have low-producing groups, leaders told they have high-producing groups will be seen by their subordinates as more apt to encourage speaking out, communicate clearly and ef-

fectively, emphasize teamwork, be open to influence, and be sensitive to differences between people.

Hypothesis 2: Subordinates in the high performance condition will have more influence during the discussion and be more satisfied with this influence than subordinates in the low performance condition. No differences are predicted for leader influence according to past performance of the group. Consistent with Tannenbaum's (1962) concept of an "expanding influence pie," it is anticipated that the leaders will maintain a relatively high degree of influence for themselves regardless of past performance, but that the subordinates will be allowed more influence when their past performance has been relatively high. Past performance will affect leadership style, which in turn will affect felt influence.

Hypothesis 3: Groups in the high performance condition will be more cohesive than groups in the low performance condition. In the high performance condition the subordinates will like each other more than in the low performance condition, and they will be less apt to want to change groups or leaders. Moreover, the leaders in the high performance condition will like their subordinates better and be less apt to want to change groups. Past performance will affect leadership style which will in turn affect members' attraction to their group.

Hypothesis 4: In the high performance condition as compared to the low performance condition, leader and subordinates will be more satisfied with each other, with the discussion, and with the solution. Subordinates will be more satisfied with their jobs and with their fellow subordinates. This greater satisfaction is anticipated as a consequence of the "better leadership," higher total influence, and greater group cohesiveness that will occur in the high performance condition.

Hypothesis 5: In the high performance condition as compared to the low performance condition, leader and subordinates will estimate greater efforts to achieve high performance and greater increases in future production. This increase in production is anticipated as a consequence of the "better leadership," higher total influence, greater cohesiveness,

and greater satisfaction that will occur in the high performance condition.

METHOD

Subjects

Two hundred persons participated in the study as members of 50 four-man groups role playing Maier's (Maier, Solem, & Maier, 1957) Change of Work Procedure case. The Ss were male graduate students in behavioral science courses at the Massachusetts Institute of Technology's Sloan School of Management and male MIT undergraduates in introductory management or behavioral science courses.³

Task

The Change of Work Procedure case involves a foreman and three workers who assemble fuel pumps in an automobile company. Maier and Hoffman (1960) describe it as follows:

The assembly operation is divided into three positions and the workers have adopted a system of hourly rotation among the three jobs. The role-playing consists of a meeting called by the foreman to discuss the possibility of their changing their work method to one in which each man works on one position only, his best position according to the time study data given to the foreman. Although theoretically the new method should increase the productivity of the workers and thus increase their piece-rate wages, the foreman's suggestion of a change to the new method usually meets with considerable resistance [p. 279].

Boredom from working on only one position is an important source of worker resistance to the suggested change.

The possible solutions to the case vary in quality and conformance to the wishes of the workers and the foreman: old (favored by the workers), new (preferred by the foreman), and integrative (an innovation solution combining positive aspects of the old and new solutions). The case has been used extensively for research purposes in the past (Hoffman, 1959; Hoffman, Harburg, & Maier, 1962; Maier, 1953; Maier & Hoffman, 1960, 1961; Maier & Solem, 1962).

Performance Manipulation

The 50 groups were randomly assigned to a high performance, low performance, or control condition by modifying the figures in the time-study report given to the foreman. In addition, the roles for the foremen in the high performance condition were modified by adding the statement: "This rate of 125% of average makes it one of the ten highest producing groups out of 50 groups in the company." In the low performance condition, foremen were

told, "This rate of 75% of average makes it 1 of the 10 lowest producing groups out of 50 groups in the company." Foremen in the control condition and workers in all three conditions received the standard role instructions (Maier et al., 1957). Twenty groups were assigned to the high performance condition, 20 to the low, and 10 to the control.

Procedure

The multiple-role-playing procedure (Maier, 1952) was used to administer the case during regular class time. The investigator read the general instructions to all groups in each class and distributed the roles to each group member, foreman and workers being assigned roles randomly. After the members had read their roles, the groups were asked to start solving the problem and to come up with a solution in 20 min. A 2-min. warning was given at the end of 18 min., and all discussion ceased at the end of 20 min. Roles were collected, and short questionnaires were administered to the foreman and three workers in each group. Each questionnaire took about 5 min. to complete.

Measurements

Perceptions of the behavior of the foreman and data on some characteristics of the decision process were obtained from each worker through the questionnaires. On his questionnaire the foreman reported the solution, perceptions of the discussion, and evaluations of the workers.⁴ Most items consisted of descriptive statements followed by 7-point scales and had been used in previous correlational studies of leadership and group behavior (Fleishman, Harris, & Burt, 1955; Likert, 1961; Stogdill, 1965; various questionnaire studies of the Institute for Social Research, University of Michigan). They are summarized in Tables 1-5.

Analysis

Several factor analyses were performed on the 18 leadership items using different samples of worker and observer data. In general these analyses supported Bowers and Seashore's (1966) four-factor theory. However, it was also possible to extract two, three, five, and six orthogonal factors,⁵ and some

⁴In 24 of the groups (12 high performance and 12 low performance), an additional student was randomly assigned to serve as an observer and complete a questionnaire virtually identical with that of the workers. Results from these untrained observers were very similar to those of the workers in describing foreman behavior, but quite different in questions that ascribed feelings to the foreman and workers. For details, see Lim (1968).

⁵When two factors were extracted, the first appeared to be a combination of interaction facilitation and support, while the second combined goal emphasis and work facilitation. When three factors were extracted, they appeared to be (a) interaction facilitation and support, (b) goal emphasis, and (c) work facilitation and close supervision.

³The authors are grateful to Thomas J. Allen, William H. Gruber, David A. Kolb, Donald G. Marquis, and Irwin M. Rubin for allowing their classes to participate in the study.

inconsistencies were found in factor structure according to the particular sample examined. For example, the item “unreasonable pressure for better performance” was more strongly associated with a support factor for one sample and a goal emphasis factor for another sample. Because of these inconsistencies it was decided to report findings for individual items, grouped by their content into Bowers and Seashore’s four factors. This grouping was carried out so as to be as consistent as possible with the results of the factor analyses that were done.

In order to test the hypotheses, *t* tests were performed comparing the high and low performance conditions. On all but three items, mentioned below, the groups in the control condition scored between the high and low groups or not significantly different from them. Therefore, their data are not shown below.

RESULTS

Validation of Experimental Manipulation

In order to determine whether the foremen responded to the information in their roles about the group’s past performance, foremen were asked to indicate after the discussion how their groups had compared to others in the company before the discussion. On a 5-point scale where 5 equals “much above average,” the foremen in the high condition scored 4.4, while those in the low condition scored 1.8, and the controls scored 3.5 (*p* < .001). Apparently the people playing the role of foreman were consciously aware of their groups’ past performance while the discussion was being conducted.

Hypothesis 1: Performance affects leadership. Subordinate perceptions of leader behavior are summarized in Table 1. Of 18 items describing leader behavior, results for 16 are in the predicted direction, and results for 11 items are statistically significant at the .05 level of confidence. Performance apparently affects a wide variety of leader behaviors. Examination of the four areas of leader behavior shows that these general findings hold for all areas, but that differences between high and low performance appear to vary according to area. Assuming that the scales are comparable, past performance appears to have its greatest effects on support and its least effects on work facilitation, with goal emphasis and interaction facilitation being about equally susceptible to influence by past performance. Leaders told that their groups are high performing are significantly more

TABLE 1
HYPOTHESIS 1: LEADER BEHAVIOR AS A FUNCTION OF PAST PERFORMANCE

| Behavior characteristic | Mean amount of behavior characteristic | |
|---|--|--------|
| | Past performance | |
| | High | Low |
| Support | | |
| Sensitive to needs and feelings of workers | 5.1 | 4.2* |
| Gives recognition for a job well done | 4.2 | 2.9*** |
| Has trust and confidence in his men | 5.1 | 4.2** |
| Punitive or critical of group’s performance | 1.8 | 3.6*** |
| Exerts unreasonable pressure for better performance | 2.8 | 3.5* |
| Goal emphasis | | |
| Lets group members know what is expected of them | 4.2 | 4.4 |
| Maintains high performance standards | 5.1 | 3.9*** |
| Stresses a feeling of pride in the group | 4.6 | 3.1*** |
| Stresses being ahead of competing work groups | 4.6 | 3.9 |
| Work facilitation | | |
| Gives reasons for suggested changes on the job | 5.3 | 4.9 |
| Allows members freedom and autonomy in their work | 4.8 | 4.1* |
| Decides in detail what shall be done and how | 2.4 | 2.8 |
| Tries to impose his preferred solution on the group | 4.2 | 4.2 |
| Interaction facilitation | | |
| Encourages speaking out and listens with respect | 5.5 | 4.9* |
| Communicates clearly and effectively | 4.8 | 4.1* |
| Emphasizes that people work together as a team | 4.0 | 3.3* |
| Open to influence from his workers | 4.8 | 4.6 |
| Sensitive to differences between people | 4.0 | 3.4 |

* *p* < .05.
** *p* < .01.
*** *p* < .001.

likely than leaders told that their groups are low performing to be seen by their subordi-

nates as sensitive, giving recognition, trusting, nonpunitive, exerting less unreasonable pressure for performance, maintaining high performance standards, stressing a feeling of pride in the group, allowing freedom, encouraging speaking out, communicating clearly, and emphasizing teamwork.

Hypothesis 2: Performance affects influence. Table 2 shows that Hypothesis 2 was strongly supported. In the high condition subordinates felt they had more influence in the discussion and were more satisfied with their influence than subordinates in the low condition. No differences were found in leader influence or satisfaction with influence according to past performance. The leaders perceived that two of the three subordinates had more influence in the high performance condition and one had more influence in the low performance condition. Apparently Tannenbaum's (1962) notion of the expanding influence pie holds in this study. With high past

TABLE 2
HYPOTHESIS 2: INFLUENCE AS A FUNCTION
OF PAST PERFORMANCE

| Measure of influence or satisfaction with influence | Mean amount of influence or satisfaction with influence | |
|---|---|-------|
| | Past performance | |
| | High | Low |
| Subordinate perception of own influence | 4.6 | 4.2* |
| Leader perception of Worker 1's influence | 4.2 | 5.4 |
| Leader perception of Worker 2's influence | 4.5 | 4.2 |
| Leader perception of Worker 3's influence | 4.8 | 4.0* |
| Subordinate satisfaction with own influence | 5.2 | 4.5** |
| Leader's perception of own influence | 4.6 | 4.8 |
| Subordinates' perception of foreman's influence | 4.1 | 3.8 |
| Leader's satisfaction with own influence | 4.1 | 4.4 |

* $p < .05$.
** $p < .01$.

TABLE 3
HYPOTHESIS 3: COHESIVENESS AS A FUNCTION
OF PAST PERFORMANCE

| Statement | Mean agreement with statement | |
|--|-------------------------------|-------|
| | Past performance | |
| | High | Low |
| By subordinates: | | |
| I like the workers in my group. | 6.2 | 5.9 |
| If same work with different group, I'd move. | 1.9 | 2.0 |
| If same work under different foreman, I'd move. | 2.5 | 3.2* |
| By leaders: | | |
| I like the men with whom I work. | 5.2 | 4.6 |
| If supervise different group, same work, I'd move. | 2.2 | 4.0** |

* $p < .05$.
** $p < .001$.

performance, subordinates' influence increased while the leader's influence remained constant.

Hypothesis 3: Performance affects cohesiveness. Table 3 shows that in the high performance condition subordinates liked each other better and wanted less to change the foremen than subordinates in the low performance condition. In neither condition were subordinates very disposed toward working with a different group of colleagues. Leaders in the high performance condition tended to like their subordinates more and were much less prone to change work groups. Apparently past performance affects attraction to a group, and especially leader-member attraction. Probably this effect of performance on cohesiveness occurs through its effect on leader behavior, which in turn affects cohesiveness.

Hypothesis 4: Performance affects satisfaction. Table 4 shows that subordinates in the high performance condition were significantly more satisfied with their fellow workers, their foreman, the discussion, and the solution than subordinates in the low performance condition. Subordinates in the high performance condition also tended to be more satisfied with their jobs. The leader was significantly more satisfied with his work group and tended to be more satisfied with the discussion and

solution in the high performance condition. Apparently past performance affects satisfaction, probably through its influence upon leader behavior.

Hypothesis 5: Performance affects future production. Table 5 shows that in the high performance condition, both the leaders and subordinates saw their groups as trying harder to achieve high performance than in the low performance condition, and the leaders in the high condition thought that their groups would maintain a higher standing in overall company performance. However, no significant differences were found according to experimental condition in changes anticipated in future production. These findings lead one to suspect that the differences obtained were due largely to the initial “set” about group performance created by the experimental instructions rather than to the discussion process itself. Had the discussion process affected feelings about future production, differences would have occurred according to experimental condition in anticipated *changes* in future production as well as in the relative standing of the groups in the company.

This interpretation is supported by a tabulation of solutions to the case produced by the groups in each experimental condition. The high performance groups produced 7 high

TABLE 5
HYPOTHESIS 5: FUTURE PRODUCTION AS A
FUNCTION OF PAST PERFORMANCE

| Estimate | Mean estimate | |
|---|------------------|-------|
| | Past performance | |
| | High | Low |
| By subordinates: | | |
| Group tries hard to achieve high performance | 5.3 | 4.6* |
| Changes in individual production | 3.5 | 3.5 |
| Changes in future production of group | 3.7 | 3.7 |
| By leader: | | |
| Group tries hard to achieve high performance | 4.5 | 2.8** |
| Future performance standing of group in company | 4.6 | 2.8** |
| Changes in future production of group | 4.0 | 3.9 |

* $p < .01$.
** $p < .001$.

quality integrative solutions and 13 lower quality old and new solutions. The low performance groups produced 10 integrative solutions and 10 lower quality solutions. Thus, no significant differences occurred in solution quality (and therefore probable future performance) according to past performance, and the tendency was for low past performance to be associated with a higher quality solution.⁶

Control Groups

In all but three instances the groups that received the standard instructions scored between the groups in the high and low conditions or not significantly different from them. These findings add support to the validity of the experimental manipulations. In both experimental conditions the leader placed more stress on being ahead of the competition than in the control condition, as one would expect. However, in the control condition the leader was more likely to decide in detail about

⁶ This tendency reached statistical significance for the first 24 groups who participated in this study (see Lim, 1968), but was reversed for the last 16 groups. The authors are currently attempting to determine reasons for these differences.

TABLE 4
HYPOTHESIS 4: SASTIFACTION AS A FUNCTION
OF PAST PERFORMANCE

| Satisfaction with | Mean amount of satisfaction | |
|-------------------|-----------------------------|--------|
| | Past performance | |
| | High | Low |
| By subordinates: | | |
| Fellow workers | 6.1 | 5.6** |
| Foreman | 5.3 | 4.6* |
| Job | 5.5 | 5.1 |
| Discussion | 4.6 | 3.8* |
| Solution | 5.5 | 4.8** |
| By leader: | | |
| Work group | 5.5 | 3.8*** |
| Discussion | 5.1 | 4.4 |
| Solution | 5.2 | 4.7 |

* $p < .05$.
** $p < .01$.
*** $p < .001$.

work activities, and the subordinates estimated that their future production would change less favorably than in either experimental condition. These differences are not readily explainable and may have been due to chance.

A Crude Reexamination of Fiedler's Theories

An important aspect of Fiedler's (1965) theories of leadership is the ability of the effective leader to be sensitive to differences between people. Two types of information were available in this study that allowed a crude test of whether this leadership trait is caused by past group performance. First a comparison was made between the high and low performance conditions in the leader's being "sensitive to differences between people" (see Table 1). Leaders who were told that they had high-performing groups were seen by their subordinates as more sensitive to differences between people, but this difference did not quite reach the .05 level of significance ($p = .06$).

Several items in the leader's questionnaire asked him to rate his three subordinates on 7-point scales on four characteristics: being an idea man, being a trouble maker, having influence in the discussion, and promotability. A tabulation was made of differences each leader saw between his subordinates on each of these scales. A comparison of leaders in the high and low performance conditions showed no differences on the average in the extent to which they saw differences between their men. A tendency occurred in only one instance for past performance to affect the leader's sensitivity of differences between his men. Leaders in the low performance condition saw greater differences between their men as trouble makers than did leaders in the high performance condition. Taken together these two analyses suggest, but certainly do not demonstrate, that a leader's sensitivity to differences between people may be in part due to the past performance of his subordinates as a group.

Comparison with Day and Hamblin's Findings

Day and Hamblin (1964) found differences in group productivity as a consequence of two dimensions of leadership that they varied ex-

perimentally: closeness and punitiveness. In the present study measurements were made of several characteristics of leader behavior dealing with closeness of supervision (e.g., unreasonable pressure, decides in detail, imposes own solution, allows freedom, encourages speaking out) and punitiveness (is punitive, sensitive to needs and feelings, gives recognition for good work, stresses group pride). Our findings indicate that performance affects leader behavior on these dimensions of closeness and punitiveness (see Table 1). Together with those of Day and Hamblin they show that performance both causes and is caused by these characteristics of leadership.

DISCUSSION

The findings of this study show that past performance affects most aspects of leader behavior, especially his support, interaction facilitation, and goal emphasis. Moreover, high past performance and the resulting leader behavior are associated with greater subordinate influence in decision making, greater group cohesiveness, and higher satisfaction. No clear relationships were found between past performance, associated leader behavior, and estimates of subsequent changes in group performance.

Unfortunately, the design of this study does not allow us to determine precisely the processes through which past performance affects leadership. Leaders given special instructions in their roles about their groups' past performance were seen by their subordinates as behaving differently according to this past performance. These differences in subordinate perception may have been due to actual differences in leader behavior. On the other hand, subordinates may have learned from their leader that they were low or high performing, attributed this past performance to his leadership capabilities, and perceived his leadership behavior during the discussion in terms of a negative or positive "halo." Reports from observers of 24 groups (data presented in Lim, 1968), which agree substantially with subordinate descriptions of leader behavior, tend to support the first interpretation. Past performance as seen by the leader clearly affects subordinate perception of

leadership and probably actual leader behavior as well.

Although these findings are based on a laboratory experiment involving role playing, two factors suggest that they may be generalized to "real world" leadership situations. First, the particular case employed was designed to simulate a real situation and has been used extensively in previous research. Second, the results of this study are consistent with those of a recent longitudinal field study by Farris (1969) who found stronger relationships between performance and organizational factors when performance was measured first.

To the extent that these findings can be generalized, they indicate that we should extend our theories of leadership and leadership training practices to account for ways in which leader behavior can (and perhaps should) occur as a consequence of past performance. Moreover, we should be especially careful in interpreting single-point-in-time correlations between leadership and performance as indicating that leadership causes performance. Clearly the causal direction can be the other way as well.

REFERENCES

- BLAKE, R. R., & MOUTON, J. S. *The managerial grid*. Houston, Texas: Gulf, 1964.
- BOWERS, D. G., & SEASHORE, S. E. Predicting organizational effectiveness with a four-factor theory of leadership. *Administrative Science Quarterly*, 1966, 11, 238-263.
- DAY, R. C., & HAMBLIN, R. L. Some effects of close and punitive styles of supervision. *American Journal of Sociology*, 1964, 69, 499-510.
- FARRIS, G. F. Organizational factors and individual performance: A longitudinal study. *Journal of Applied Psychology*, 1969, 53, 87-92.
- FIEDLER, F. E. The contingency model: A theory of leadership effectiveness. In H. Proshansky & B. Seidenberg (Eds.), *Basic studies in social psychology*. New York: Holt, Rinehart & Winston, 1965.
- FLEISHMAN, E. A., HARRIS, E. F., & BURTT, H. E. *Leadership and supervision in industry*. Columbus: Bureau of Business Research, Ohio State University, 1955.
- HOFFMAN, L. R. Homogeneity of member personality and its effect on group problem-solving. *Journal of Applied Psychology*, 1959, 58, 27-32.
- HOFFMAN, L. R., HARBURG, E., & MAIER, N. R. F. Differences and disagreement as factors in creative group problem solving. *Journal of Abnormal and Social Psychology*, 1962, 64, 206-214.
- JACKSON, J. M. The effect of changing the leadership of small work groups. *Human Relations*, 1953, 6, 25-44.
- KATZ, D., & KAHN, R. L. Some recent findings in human relations research in industry. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1952.
- KATZ, D., & KAHN, R. L. Leadership practices in relation to productivity and morale. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory*. (2nd ed.) Evanston, Ill.: Row, Peterson, 1960.
- KATZ, D., MACCOBY, N., GURIN, G., & FLOOR, L. G. *Productivity, supervision and morale among railroad workers*. Ann Arbor: Institute for Social Research, University of Michigan, 1951.
- KATZ, D., MACCOBY, N., & MORSE, N. *Productivity, supervision and morale in an office situation*. Ann Arbor: Institute for Social Research, University of Michigan, 1950.
- LICKERT, R. *New patterns of management*. New York: McGraw-Hill, 1961.
- LIM, F. G., JR. Some effects of performance upon leadership behavior. Unpublished master's thesis, Sloan School of Management, Massachusetts Institute of Technology, 1968.
- MAIER, N. R. F. *Principles of human relations*. New York: Wiley, 1952.
- MAIER, N. R. F. An experimental test of the effect of training on discussion leadership. *Human Relations*, 1953, 6, 161-173.
- MAIER, N. R. F., & HOFFMAN, L. R. Quality of first and second solutions in group problem solving. *Journal of Applied Psychology*, 1960, 44, 278-283.
- MAIER, N. R. F., & HOFFMAN, L. R. Organization and creative problem solving. *Journal of Applied Psychology*, 1961, 45, 277-280.
- MAIER, N. R. F., & SOLEM, A. R. Improving solutions by turning choice situations into problems. *Personnel Psychology*, 1962, 15, 151-157.
- MAIER, N. R. F., SOLEM, A. R., & MAIER, A. *Supervisory and executive development: A manual for role playing*. New York: Wiley, 1957.
- MCGREGOR, D. *The human side of enterprise*. New York: McGraw-Hill, 1960.
- STODILL, R. M. *Managers, employees, organizations*. Columbus: Bureau of Business Research, Ohio State University, 1965.
- TANNENBAUM, A. S. Control in organizations: Individual adjustment and organizational performance. *Administrative Science Quarterly*, 1962, 7, 236-257.

(Received December 16, 1968)

SKIMMING LISTS OF FOOD INGREDIENTS PRINTED IN DIFFERENT BRIGHTNESS CONTRASTS¹

E. C. POULTON²

Applied Psychology Research Unit, Cambridge, England

Seventy-six people aged 29-72 yr. searched for particular words in lists of ingredients printed on white paper, reflectance 85%, in ink with densities of 1.3, .4, .2, and .1 (reflectances 4, 34, 53, and 68%). The print was 6-pt. lower-case Univers with .6-pt. leading. There were four sets each of 15 lists. A 4×4 factorial design was used that confounded list difficulty with order. In separate experiments the lighting was 40 and 2 ftc. There were large drops ($p < .01$) in the rate of locating ingredients when the density of the ink decreased from .4 to .1. Increasing the density from .4 to 1.3 had no reliable effect ($p > .05$). Two people failed to locate any ingredients in the poor light when the density of the ink was .1. It was concluded that ingredients printed in 6-pt. lower-case Univers on white paper should have an ink density of at least .4. The contrast ratio between the ink and the paper is then at least 60%; the relative brightness ratio is at least 2.5 : 1.

Designers of packages sometimes use pale letters on pale backgrounds in order to portray softness or femininity. For other packages they may use dark letters on dark backgrounds. Some of the recent covers of the *Journal of Applied Psychology* and of *Human Factors* fall into this category. It is clear from the threshold determinations of Cobb and Moss (1928) that poor contrast between letters and background makes discrimination more difficult. Either the critical details need to be larger, or else more light is required.

The present experiment is the second in a series. In the previous article (Poulton, 1969) it was concluded that lists of food ingredients should not be printed in Univers smaller than 6 pt. The present experiment used 6-pt. Univers printed on white paper with ink of various densities. The aim was to determine the lowest density of ink that did not appreciably retard the housewife when she skimmed lists of food ingredients. The same two levels of lighting were used as in the

previous experiment. They are intended to represent, respectively, the lighting in a supermarket and in a domestic kitchen cupboard.

Tinker and Paterson (1931) compared various colors of print and background. They used their speed of reading method. They concluded that speed depended upon the brightness contrast between the print and the background. Unfortunately they did not specify the brightness contrasts that they used.

Other methods have been used to compare print and backgrounds with various brightness contrasts and color contrasts. They are not appropriate to the problems of the housewife who is skimming a list of food ingredients printed on a package that she has picked up from a shelf. One method is to compare the distances at which the print can just be read. Another method is to compare the amount of print that can be read in a brief glance, using a tachistoscope. The results have been summarized by Tinker (1963, pp. 137-148).

Williams (1967) recently used a related method. Four 3-digit numbers were presented at a time. The digits were 9 pt. (3 mm. tall). Eighteen men had to search for a particular 3-digit number, and indicate its position as quickly as possible by pressing one of four buttons. Reaction time was the chief measure of visibility. Williams varied the brightness

¹ This research was carried out at the suggestion of The Metal Box Co. Ltd., which supplied the printed materials. The British Food Manufacturers Federation kindly defrayed the cost of the Ss. The author is grateful to I. Harris of The Metal Box Co. Ltd. for his help and encouragement. Financial support from the British Medical Research Council is also gratefully acknowledged.

² Requests for reprints should be sent to E. C. Poulton, Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

contrast and the lighting, but he did not reduce the contrast ratio below 50%. Also he varied the brightness contrast by changing the reflectance of the background. This changes the amount of light reflected by the display, and is equivalent to varying the lighting (Cobb & Moss, 1928). The effects of brightness contrast are therefore confounded with the effects of the amount of light reflected into the eyes.

There were no reliable effects on reaction time with an illumination of 6 ftc. With an illumination of .6 ftc. there was a reliable gap of over 1.0 sec. between the worst condition and any of the other conditions. The worst condition was black numbers on a background with a reflectance of 8%. This gives a contrast ratio of $8 - 4/8$ or 50%. After the gap the next worst condition was black numbers on a background with a reflectance of 16%. This gives a contrast ratio of $16 - 4/16$ or 75%, and about twice as much reflected light. Black numbers on a white background had the shortest reaction times. This gives a contrast ratio of $83 - 4/83$ or 95%, and about 10 times as much reflected light as the worst condition. The remaining six conditions all fall in the gap of .5 sec. between these two. Unfortunately the results are due partly to differences in the amount of reflected light. They are determined only partly by differences in the brightness contrast.

METHOD

Materials

The same lists of ingredients were used as in the previous experiments (Poulton, 1969). All four sets were reproduced photographically in 6-pt. Univers with .6-pt. leading between lines. They were printed four times in inks with densities of 1.3, .4, .2, and .1 on white paper. The paper was Spartocote. It had a reflectance of about 85% and a slight shine. The reflectances of the inks were 4, 34, 53, and 68%, respectively.

Experimental Design and Subjects

Two experiments were run with different levels of lighting. A total of 36 Ss worked with about 40 ftc. on the printed pages. Another 40 Ss worked with about 2 ftc. In each experiment the four densities of ink and four sets were arranged in a Latin-square design with four groups of Ss. The lists were presented always in the same order, as in

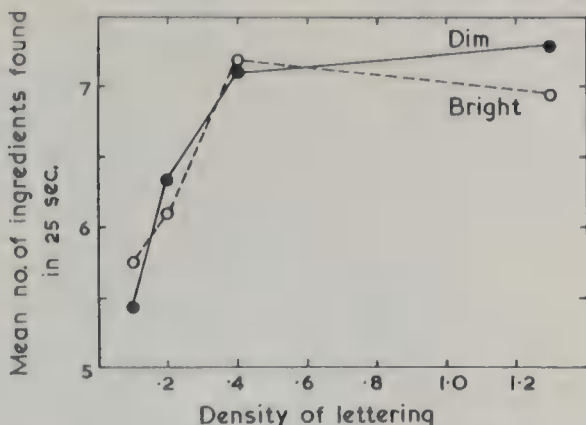


FIG. 1. Average number of ingredients found in 25 sec. with various densities of ink. (Unfilled points: about 40 ftc. on table top; data from 36 Ss. Filled points: about 2 ftc. on table top; data from a separate group of 40 Ss.)

the previous experiments. The Ss were randomly allocated to groups.

The 76 Ss were members of a panel maintained at the Applied Psychology Research Unit at Cambridge, England. In the experiment with the bright light the ages ranged from 29 to 72 yr. One-third were men; the remainder were women. In the experiment with the dim light the ages ranged from 29 to 58 yr. There was only one man. About two-thirds wore glasses for reading. Two others said that they should have done so. They were paid \$.90 (7 shillings and 6 pence) per hour for their services, plus traveling expenses.

Procedure

The procedure followed that of the previous experiments. The Ss were tested in groups, seated at tables. There were two practices to show Ss what to do. The lists used in the second practice were printed in inks of the same density as the first test lists. The S had to read a target word on a question sheet, find the word in the corresponding list of ingredients, and cross it out. There were 15 target words, 1 in each of 15 lists. As many target words as possible had to be crossed out in the 25 sec. available. The practices and the four parts of the experiment together took about 15 min.

RESULTS AND DISCUSSION

Figure 1 shows the average number of ingredients found in 25 sec. The data have been pooled over all Ss. Analysis of variance followed by Tukey's range test (Ryan, 1959, Appendix) was carried out separately for each level of lighting. In the bright light, ink densities of .4 and 1.3 both produced reliably quicker work than ink densities of either .1 or .2 ($p < .05$ or better).

In the dim light, the ink density of .2 produced reliably quicker work than the ink density of .1 ($p < .05$). The ink density of 1.3 produced reliably quicker work still ($p < .05$). Two housewives failed to locate any ingredients printed with ink densities of .1. Both were over 40 yr. old.

The results in Figure 1 suggest that for the 6-pt. Univers used, the ink density should not be less than .4 when printing on white paper. With densities less than .4 there is a sharp fall in visibility as measured by the rate of locating ingredients. Densities greater than .4 make little difference to the rate. An ink density of .4 corresponds to a contrast ratio between the lettering and the white paper of $85 - 34/85$ or 60%. The relative brightness ratio is $85:34$, or 2.5:1.

The results are not really comparable to the results of Williams (1967) that were referred to in the introduction. Williams found no reliable differences with an illumination of 6 ftc. This lies between the illumination levels of about 40 and 2 ftc. used here, both of which gave reliable differences when the brightness contrast was varied. Figure 1 shows

that the rate of work did not fall appreciably until the density of the lettering had been reduced to .2. This gives a contrast ratio of only $85 - 53/85$ or 38%. Williams did not include any contrast ratios as low as this in his experiment. Thus the lighting had to be reduced to .6 ftc. even with dark backgrounds in order to produce reliable differences between his experimental conditions.

REFERENCES

- COBB, P. W., & MOSS, F. K. The four variables of the visual threshold. *Journal of the Franklin Institute*, 1928, **205**, 831-847.
- POULTON, E. C. Skimming lists of food ingredients printed in different sizes. *Journal of Applied Psychology*, 1969, **53**, 55-58.
- RYAN, T. A. Multiple comparisons in psychological research. *Psychological Bulletin*, 1959, **56**, 26-47.
- TINKER, M. A. Legibility of print. Ames: Iowa State University Press, 1963.
- TINKER, M. A., & PATERSON, D. G. Studies of typographical factors influencing speed of reading: VII. Variations in color of print and background. *Journal of Applied Psychology*, 1931, **15**, 471-479.
- WILLIAMS, C. M. Legibility of numbers as a function of contrast and illumination. *Human Factors*, 1967, **9**, 455-460.

(Received January 2, 1969)

EFFECT OF INTERVIEWS ON TEACHER SELECTION DECISIONS¹

DALE L. BOLTON² AND MICHAEL E. HICKEY

University of Washington

An experiment was conducted in a simulated situation in which 144 elementary school principals made decisions regarding fictitious applicants for a hypothetical position. The purpose of the experiment was to determine how interview information affects (a) time required to make decisions, (b) feeling of certainty regarding decisions, (c) fineness of discriminations made, and (d) consistency of decisions. The levels of the interview information were (a) audiovisual (via color films), (b) audio (sound track of the films), and (c) no interview information. Interviews were scripted and role played to control content. The results indicated that interview information increased discrimination and time but had no effect on consistency. More certainty resulted when both audio and visual stimuli were used; otherwise, audio interview information had the same effect as audiovisual information.

Interviews have been used almost universally as a part of the teacher selection process for many years. Although widespread use of the interview indicates a certain belief in its utility, the actual value of the interview can be ascertained in only three ways.

First, the value of the interview can be determined by the extent to which it helps to predict which teachers will be most successful. The predictive validity of the interview has been the basis of much research. Interviews are not generally predictive; that is, they are not generally valid. Rather, their validity must be determined in a given situation, for particular positions, and following specified procedures. Because of the situational aspect of establishing the predictive validity of interviews, the question of the "goodness" of a decision in terms of whether the "correct" teacher was selected was intentionally omitted from this study. It is assumed that local school systems define teacher effectiveness according to specified local criteria; if so, the local system will be able to specify the outcomes desired in terms of teacher behavior.³

Second, an indication of the value of the interview may be obtained by determining how the interview contributes to factors necessary for successful prediction. For example, unless measures obtained in the interview are both *discriminating* and *consistent*, predictions made from these measures will be meaningless or spurious. A nondiscriminating measure approaches a constant; therefore, any correlation with the measure will approach zero. A nonconsistent measure is likely to yield a high correlation one time and a low or negative correlation another time. Therefore, discrimination and consistency are necessary conditions for predictive validity, but they are not sufficient. The study reported here answers the question of whether interviews increase discrimination and consistency in selection decisions, but the limits of the study do not allow conclusions about predictive validity.

This study extends the work of others who have examined the interview's utility along dimensions other than predictive validity. Sydiaha (1961, 1962) studied the effect of an interviewer's empathy on the accept-reject decision. Springbett (1958) and Bolster and Springbett (1961) studied the effect that appearance of the applicant and order of presentation of interview information made on decisions. Levine and McGuire (1968)

cants in a simulated situation can assess teacher applicants consistently according to selection criteria specified by a local school district. This assumption should be verified empirically.

¹ This study was partially supported by a grant from the United States Office of Education, Grant No. OEC 4-7-061349-0266, *Variables Affecting Decision Making in the Selection of Teachers*. Director: Dale L. Bolton, August 1968.

² Requests for reprints should be sent to Dale L. Bolton, College of Education, University of Washington, Seattle, Washington 98105.

³ Also, it is assumed that the decision maker who can discriminate consistently among teacher appli-

also studied the order effect of interview information, noting that early cues tend to distort medical diagnoses. Barrett (1958) found that psychologists who were required to write evaluative reports on patients tended to weight interview information heavier than objective data.

In Giedt's (1955) study of the accuracy of personality judgments by clinical psychologists, Ss were presented information about patients via visual, verbal, verbal plus sound, or audiovisual means. In this setting, Giedt concluded that (a) evaluations and predictions are separate unrelated tasks of the decision process and that cues affect the two tasks differentially; (b) as long as content is available, no differences occur in diagnosis; and (c) interviewers should be careful to avoid being misled by the patient's appearance and expressive cues. Giedt's study suggested the need to measure predictions or estimations of consequences as well as relative evaluations of applicants. Likewise, Giedt suggested the need to manipulate audio and visual cues rather than study the interview process in its natural setting as Sydiaha did in his 1961 study.

The third indication of the value of the interview may be obtained by determining whether it contributes to the efficiency of the selection process. Two clues to efficiency are the amount of *time* needed to make decisions and the confidence or *certainty* with which the decision maker regards his decisions. The importance of the time factor seems obvious in that a small amount of time saved on each of a large number of teacher selection decisions means a considerable saving in time and money to the school district. The significance of the certainty factor is based on the view that decisiveness in an administrator is a good quality, that uncertainty can lead to indecision, and that indecision can cause vacillation and wasted motion. This study answers the question of how interview information affects time and certainty in teacher selection decisions.

The value of the teacher selection interview was investigated by determining the contribution of interview information to (a) *discrimination* and *consistency* of decisions (two factors necessary to predictive validity) and

(b) *time* needed to make decisions and feeling of *certainty* regarding the decisions (two factors related to the efficiency of the selection process). The study was designed so that the effects of the audio and visual stimuli of the interview could be determined on these four dependent variables.

PROCEDURE

The Ss of the experiment were 144 elementary school principals, many of whom were involved in the selection process on a seasonal basis. The Ss were randomly divided into three equal-sized groups⁴ for receiving interview information. One group received audiovisual information, another group received audio information only, and a third group did not receive interview information. In order to control the presentation of the interviews for the experiment, it was not possible to use "live" interviews in which Ss actually interviewed applicants.⁵ Consequently, interviews were scripted and role played to control content and filmed to control the presentation. The sound and color film of each applicant was used for audiovisual treatment, while only the tape-recorded sound track of this film was used for the audio treatment. The camera focused on the applicant throughout the entire interview, and the interviewer remained anonymous.

The Ss were oriented to a simulated teacher selection situation in which they considered eight fictitious applicants for a well-described, hypothetical, fourth grade teaching position. All Ss were provided written documents on each applicant. The documents were similar to those used by personnel directors, including information from letters of application, credentials, and recommendations.

The general experimental task performed by each S was to make decisions about the appropriateness of each applicant for the position on the basis of the information available. Each S was asked to (a) estimate how each applicant would be evaluated on a Teacher Evaluation Instrument (TEI) at the end of 1 yr. of teaching, (b) rank the eight applicants according to their suitability for the hypothetical situation, and (c) indicate the degree of certainty of his judgments regarding the estimates on the TEI and ranking by indicating how willing he would be to bet that his judgments were correct.

The above tasks were completed during the morning session of the experiment. For purposes of measuring the consistency of the decisions, a retest was administered in the afternoon in the following manner. Five of the eight applicants presented in the first session were repeated in the afternoon

⁴ Three other variables that were manipulated in the study are not discussed in this article. They were number of documents, masking of information, and instructions regarding the processing of information.

⁵ In effect, the interviewing skill of S was not included in the experiment—his use of information from the interview was measured, however.

session. These five applicants were made to appear different by modifying certain minor data in their records, for example, age, birthplace, height, and weight. Changes in make-up, hairpieces, and clothes altered appearances during the filmed interview. The other three applicants used during the morning session were decoys and were replaced by considerably different applicants in the afternoon session. The decoys appeared late in the order of presentation in the first session and early in the second session to aid in forming the impression that the second set was an entirely new set of applicants. It was assumed that the insertion of the decoys did not affect decisions about the five applicants on whom repeated measures were taken.

When these tasks had been completed, it was possible to measure the four dependent variables in the following way:

1. *Time* was measured directly by the number of minutes required to complete the total task.

2. *Certainty* was measured directly according to S's willingness to bet that his judgments were correct. Two expressions were obtained (regarding judgments on the TEI and the ranking), and separate analyses were made of each.

3. *Discrimination* was determined by computing the variance of the 16 applicant scores on each item of the TEI; the mean variance of all items was then used as a measure of discrimination. The greater the variance, the more discriminating the individual; the smaller the variance, the less discriminating.

4. Two measures of *consistency* were obtained: (a) a correlation between the first and second ranking of the five "real" applicants, that is, omitting the three decoys from the morning and afternoon sessions; and (b) a correlation between the first and second estimates of how each of the five applicants would be evaluated on the TEI. The correlations were transformed by Fisher's r to z transformation and the z scores used in separate analyses.

Development of the Simulated Interviews

The development of the filmed interviews was subject to a number of constraints. First, because of time and fiscal limitations, the length of each interview was limited to 9 min. This made it necessary to display only the *probing* portion of the interview, which was considered to be the most vital part. Second, all applicants for the position were assumed to be at least minimally qualified for employment, and all had presumably passed an initial screening interview although none of this initial interview information was available to Ss. Third, as a control measure, all applicants were female, all were between 22 and 28 yr. old, and all were of acceptable appearance; that is, none was at either extreme in terms of physical appearance. The controls provided a group of applicants that were relatively homogeneous with respect to these classification variables; relative homogeneity was necessary for testing discrimination among treatments.

The problem confronted in preparing these materials was how to display the personality character-

istics of the applicants in a 9-min. segment of the interview so as to permit them to be assessed and rated by Ss. Five factors were delineated by Ryans (1960) as significant in describing teacher behavior for the elementary teacher group in his study. These factors (described in detail in Ryans' work but identified here as buoyancy, empathy, organization, originality, and sociability) were manipulated among the 16 fictitious applicants in such a way that a personality was "created" in which each applicant was obviously high on one factor (e.g., originality); two other factors were less obvious, but present in the interview behavior (e.g., buoyancy and sociability); and two other factors were not evident (e.g., organization and empathy).

Once the personalities of the applicants had been created, it was necessary to prepare scripts that would display these characteristics in natural responses during the filmed interview. Analytical and probing questions that suggested an extended answer and that might reasonably be asked in an interview were devised. Scripts were then written for each applicant in order to control the time element and the predetermined characteristics of the applicants.

The individuals used to portray the fictitious applicants in the filmed interviews were selected from senior University of Washington education majors. The actresses were selected on the basis of the extent to which they "fit" one of the fictitious applicants whose traits they would display on the filmed interviews. The scripts were memorized and rehearsed until a natural aura pervaded the interview. Information in the documents given to Ss was complementary to the interview information in that it portrayed the same characteristics.

The rationale for the design of the interview was to display specific behavioral and personality factors that could be assessed by Ss. Because the study was designed to determine Ss' certainty of assessment of the applicants and their ability to discriminate among the applicants, the structure of each interview was developed so as to control stringently the characteristics displayed by the applicant in the interview, while retaining the realism and spontaneity of the situation as far as possible.

Subsequent use of the filmed interviews in the research project and related presentations obtained an overwhelmingly favorable response to their realism. As was expected, however, a generally negative response was elicited concerning the narrow focus of the interview segment; some reactors thought it failed to display enough of the "total personality" of the applicant, despite the fact that all of the major dimensions to be assessed in the experiment were evident in the interview and repeated in related documents.

RESULTS

Separate analyses of variance were made for the effect of the interview information on time, discrimination, certainty, and consistency. The analyses indicated that there were

TABLE 1
EFFECT OF VARIOUS LEVELS OF INTERVIEW INFORMATION ON TIME, DISCRIMINATION,
CERTAINTY, AND CONSISTENCY

| Dependent measure | Level of interview information ^a | | | <i>MS_b</i> | <i>MS_w</i> | <i>F^b</i> |
|-------------------|---|-------|-------|-----------------------|-----------------------|----------------------|
| | Audio-visual | Audio | None | | | |
| Time | 274.1 | 283.8 | 219.1 | 58392.1 | 1317.8 | 44.3** |
| Discrimination | 1.150 | 1.188 | .925 | .969 | .190 | 5.10** |
| Certainty | | | | | | |
| On TEI | 9.625 | 8.646 | 8.958 | 12.007 | 2.887 | 4.16* |
| On ranking | 9.917 | 8.500 | 9.042 | 24.528 | 4.074 | 6.02** |
| Consistency | | | | | | |
| On TEI | .371 | .369 | .428 | .055 | .059 | — |
| On ranking | .549 | .628 | .750 | .489 | .359 | 1.36 |

Note.—TEI = Teacher Evaluation Instrument.
^a Entries are mean scores.
^b *df* = 2/108.
* *p* < .05.
** *p* < .01.

effects on time, discrimination, and certainty; but whether an *S* received no interview information, audio interview information, or audiovisual interview information had no effect on the consistency with which he ranked the applicants or estimated consequences on the Teacher Evaluation Instrument (TEI). These results are shown in Table 1.

Interview information affected the time required to make decisions in the following way. Audiovisual and audio information were not significantly different,⁶ but both required *more time* to reach a decision than no interview information. This was a natural consequence of the time needed to obtain the information from the interview and was anticipated as a result of the differences in treatment.

With relation to discrimination on estimating ratings on the TEI, the results of audiovisual and audio interview information were not significantly different, but in each case the results were *more discriminating* than no interview information. This result indicates that information was obtained from the content of the interview that was not obtained from the written documents but that seeing the applicant did not make *S* more discriminating than listening to the applicant.

The *Ss* were required to estimate their feel-

ing of certainty regarding two tasks: their estimates of ratings on the TEI, and their ranking of the applicants. On the estimates of consequences on the TEI, there was significantly more certainty expressed with audiovisual than with audio interview information. No interview information yielded a mean score that was not significantly different from the other scores. On the feelings of certainty about ranking of applicants, audiovisual interview information yielded significantly more certainty than either audio information or no interview information. For each estimate of certainty (i.e., ratings on the TEI, and rankings of applicants), there was more certainty expressed when audiovisual information was received than when only audio information was received. Even though this increased certainty was expressed, the decisions were no more discriminating, no more consistent, and took no less time than when only the audio information was received. This suggests that seeing an applicant in an interview does not affect the decision itself as much as it does the confidence of the decision maker.

The results of the study do not answer questions about whether the nature of interview information affects the types of cues attended to (as suggested by Giedt; Levine & McGuire; Sydiaha; and the Springbett

⁶ A Neuman-Keuls test was used for all post-analyses of means, and the significance level required was at least .05.

studies), but they do extend the prior findings by suggesting that

1. Consistency of ranking of applicants or of estimating consequences is not significantly affected by interview information. It was anticipated that the information included in the interview (both the content and the expressive cues) would increase consistency, but such was not the case. It may have been that the compatibility of the information in the interview with the information in the documents contributed to these results.

2. Interview information did increase the discrimination of Ss, as anticipated, but seeing applicants was no more helpful than hearing them. These results are different from the results of prior studies which suggest that expressive cues adversely affect decisions; they appear to be more compatible with Giedt's results where content of the interview was the predominant influence on diagnostic decisions. However, it should be remembered that prior studies were not concerned with the measurement of discrimination.

3. Seeing applicants increased Ss' feelings of certainty about applicants. This result, too, is compatible with Giedt's findings, but extends them by indicating that while the judgments made in the selection process may not be affected by visual cues, the feeling of certainty regarding these judgments may be increased. Certainty under these conditions may be a function of a lack of experience with audio stimuli only and may be amen-

able to training. If so, there would be no apparent disadvantage to audio interviews, for example, via telephone.

4. As anticipated, total time taken for the decision-making task was greater when interview information was received. It was thought that the uncertainty accompanying no interview information might cause sufficient vacillation to increase the time; but if this increase occurred, it was insufficient to offset the time spent with the interview information.

REFERENCES

- BARRETT, R. S. The process of predicting job performance. *Personnel Psychology*, 1958, 11, 39-55.
- BOLSTER, B. I., & SPRINGBETT, B. M. The reaction of interviewers to favorable and unfavorable information. *Journal of Applied Psychology*, 1961, 45, 97-103.
- GIEDT, H. F. Comparison of visual, content, and auditory cues in interviewing. *Journal of Consulting Psychology*, 1955, 19, 407-416.
- LEVINE, H. G., & MCGUIRE, C. Role playing as an evaluative technique. *Journal of Educational Measurement*, 1968, 5, 1-8.
- RYANS, D. G. *Characteristics of teachers*. Washington, D. C.: American Council on Education, 1960.
- SPRINGBETT, B. M. Factors affecting the final decision in the employment interview. *Canadian Journal of Psychology*, 1958, 12, 13-22.
- SYDIAHA, D. Bales' interaction process analysis of personnel selection interviews. *Journal of Applied Psychology*, 1961, 45, 393-401.
- SYDIAHA, D. Interviewer consistency in the use of empathic models in personnel selection. *Journal of Applied Psychology*, 1962, 46, 344-349.

(Received January 2, 1969)

ON THE USE OF ALTERNATIVE ORGANISMS

DOUGLAS A. BERNSTEIN¹

AND

THOMAS M. ALLOWAY

University of Illinois

University of Toronto

The use of nonhuman organisms to perform human tasks is discussed. It is pointed out that application of animal behavior to human activity is far from new. Recent "crackpot" ideas are reviewed and suggestions for judicious extension of the principle are made.

In spite of the technological explosion in the West, a good deal of human activity still involves arduous, monotonous, and even degrading physical labor and, sometimes, considerable danger. The most obvious examples of such activity include jobs such as painting and maintenance of tall buildings and bridges, fruit picking, bomb disposal, garbage collection, mass production assembly, etc. It should be made clear that, in many cases, man is engaging in activities to which he, as an organism on the phylogenetic scale, is not ideally suited and that could be performed more efficiently and profitably by alternative organisms.

This is by no means a new idea. Bernstein (1965) pointed out that "Whenever human beings are paid a wage for the use of their sense organs rather than for intelligence or judgment, it is likely that they could be replaced by animals, economically and easily [p. 103]." The problem, as Cumming (1966) noted is that we sometimes have difficulty "telling when the unique capacities of the human organisms are largely wasted in trivial performances that lower organisms are perfectly capable of mastering and better able than we to tolerate [pp. 246-247]."

The purpose of the present article is to demonstrate that substitution of nonhuman organisms in the performance of typically human tasks has already shown its feasibility and usefulness and to suggest in addition that systematic expansion of such substitution is desirable.

Existing Data on the Use of Alternative Organisms

It hardly seems necessary to document the role of nonhuman organisms in the life of man.

¹ Requests for reprints should be sent to Douglas A. Bernstein, University of Illinois, Psychological Clinic, Children's Research Center Building, Champaign, Illinois 61820.

Animals have been employed for thousands of years in a wide variety of ways, most of which involve simple exploitation of their already available response repertoires. Dogs herd cattle and sheep, hunt and track, guard property, and lead the blind. Pigeons carry messages; horses and other large animals aid in agricultural activities and provide transportation, while nearly every kind of nonhuman from flea to elephant acts as entertainers (e.g., Breland & Breland, 1951, 1961). The list can be lengthened almost endlessly, though this need not be done to illustrate the point that men and animals have led, and continue to lead, interacting lives.

With the advent of World War II, the range of tasks for which animals are used began to expand, though again basic response repertoires were exploited, not modified. As one might expect, most of the new tasks were, at first, military. For example, Skinner (1956) noted that dogs were trained by the Russians to disable tanks by running close alongside with magnetic mines, and that the British trained sea gulls as submarine detectors (Skinner, 1960). In the latter case, British submarines were sent out along the coast to release food while running submerged. The birds soon learned to follow submerged vessels and were never taught a German-British discrimination.

Skinner (1960) also mentioned (though did not vouch for) a report of the Russians' use of sea lions to sever cables attached to floating mines. The procedure apparently involved fitting the animals with an electrical cutting device that, when correctly brought into the proximity of a cable, both closed the blades and dispensed a few fish from a small tank. When battery power became insufficient to drive the blades, a discriminative stimulus sent the animals back to base for resupply.

Perhaps the most famous wartime use of alternative organisms was in the pigeon guidance system, designed and perfected by Skinner, which allowed a winged aircraft, loaded with explosives, to be brought in on a designated target by a team of on-board pigeons that pecked at a visual target display. The details of the arrangement are available elsewhere (Skinner, 1960). Even though the system worked with great accuracy and was demonstrated to be feasible, it was rejected by a group of scientists because "the spectacle of a living pigeon carrying out its assignment, no matter how beautifully, simply reminded the committee of how utterly fantastic our proposal was [Skinner, 1960, p. 34]."

Another rejected wartime plan suggested dropping thousands of "incendiary bats" on enemy cities. Each bat would have carried an incendiary time bomb, all of which would have exploded simultaneously after the bats had settled under eaves, in attics, and elsewhere (Skinner, 1960). In spite of the rather bleak reception given such imaginative proposals in the past, the expanded use of animals in military settings is still a very real, though not necessarily desirable, possibility. It already has been demonstrated, for example, that pigeons might be used to analyze photo reconnaissance material (Herrnstein & Loveland, 1966), and there is apparently some classified research underway relating to the military uses of pigeons (Herrnstein, 1965).

Nonmilitary uses of alternative organisms in which no significant repertoire modifications were made have been reported also. Probably the most significant of these has been in the area of quality control in industrial settings. Verhave (1966), working at a large pharmaceutical company, trained pigeons to inspect gelatin capsules for defects. The birds did so by pecking one of two keys. In no case did birds approach or make contact with the capsules, and their performance after a week of training had reached 99% correct. However, the system was rejected by the company, apparently because it feared adverse consequences for public relations.

This project, while it lasted, stimulated Cumming (1966) to develop a similar procedure at another industrial firm where, in this case, the pigeons were trained to inspect

diodes for paint defects. The results gave every indication that the birds were capable of levels of speed, accuracy, and endurance that surpassed those of human inspectors, but again, the idea of incorporating the system into the production line of the company was rejected, this time due to questions about the response of animal lovers and, perhaps more realistically, organized labor.

So far we have been talking about using alternative organisms whose behavioral repertoires are left essentially intact. There are, however, many examples in which significant repertoire modifications have been accomplished, that is, cases in which animals are trained not simply to, say, peck at a key rather than at grain, but to emit new responses not seen in the naturally occurring behavior of the organism. Such modifications are most evident in cases where animals are trained as entertainers. Breland and Breland (1951, 1961) report "acts" involving pigs that operate vacuum cleaners, chickens that bat and retrieve balls, raccoons that deposit coins in a bank, etc., and it is easy for all of us to recall having seen countless other animal entertainers who behave in similarly "non-animal" ways.

But perhaps the most exciting examples in which behavior repertoires are modified involve the primates, for here we have an alternative organism whose physical structure (and thus, potentially, whose behavior) most closely approximates that of man himself. "Monkey acts" have long been among the biggest attractions at zoos, circuses, and in Tarzan movies, partly because these animals can do so many "near-human" things. It has not been until recently, however, that the potential usefulness of primates as alternative organisms for nonentertainment purposes has begun to gain recognition.

Not all of the applications have been legal. Clarke (1958) reported a case on the New York City police blotter in which a chimpanzee had been trained to burglarize upper-floor apartments by climbing the side of the building and entering through a window.

The most spectacular case, which is really no more than a monkey act that got out of hand, was reported on a short-wave radio broadcast from Johannesburg, South Africa.

Apparently, South Africans have trained baboons to drive tractors and are employing them in the planting, cultivating, and harvesting of crops in that country. (An inquiry, requesting further details and references, was sent to Radio South Africa, but was ignored).

The preceding series of what are, to us, impressive examples of the use of alternative organisms has been presented in order to make one major point: A good deal of work has now been done, both in earnest and in jest, which, when viewed in light of an ever evolving external control technology begun by Skinner (as well as of an internal control technology, e.g., Delgado, 1963, 1965), clearly demonstrates the potential for employing nonhumans in a range of activities that is limited only by the bounds of our ingenuity and the presence and nature of species-specific characteristics.

Suggested Applications and Extensions

Given the existing data and the presence of the technology, it does not seem unreasonable to make a series of suggestions for ways in which alternative organisms might be employed in the future.

The most obvious areas of attention are those in which some evidence of success is already available. Thus, it is suggested that efforts be made to expand the use of alternative organisms of various kinds in agricultural settings (e.g., primates as fruit pickers, tractor operators, general maintenance workers, etc.) as well as in quality control (pigeons are the obvious choice here) and certain types of assembly work (primates would probably do very well at this).

Moving into hitherto unexplored areas,² it seems perfectly sensible to investigate the feasibility of training baboons or other primates to act as garbage or litter collectors, dock workers, window washers, street sweepers, and bellboys, to name a few possibilities. Because of their behavioral potential (and their expendability), primates might also make excellent astronauts on certain types of missions.

A variety of animals might be employable

² Some of the following suggestions (and some others that do not appear here) have also been made by Clarke (1958).

as rural (or even urban) mailmen. Porpoises could be trained systematically to cruise as resident life guards off beaches and lake shores (it has been reported anecdotally that these animals have saved swimmers' lives even without such training), and pigeons, it would seem, would make excellent radar screen observers, especially where constant surveillance is required in areas (such as along the DEW line) where the climate makes the continued presence of human personnel absurd.

These suggestions are meant only to stimulate further thought and additional ideas.³ They are offered in the hope that it may some day be possible to relieve man of "some of the more odious . . . tasks on which the capabilities of human beings for extremely complex judgments and decisions are wasted [Ulrich, Stachnik, & Mabry, 1966, p. 238]."

Readers unacquainted with the literature on animal training may be understandably skeptical about the possibility of training animals to perform tasks such as those proposed here. The techniques for doing so, however, are actually rather straightforward. In many instances, the first step would be to place the animal on a "token reward system" in which correct behavior is maintained by giving the animal a poker chip or other easily dispensed token, which he may spend later to obtain the necessities and luxuries of life, such as food or his mate. The procedure for establishing the value of tokens is a simple one (Wolfe, 1936) in which the animal is trained to insert the tokens into a vending machine to obtain food. Once the animal has learned to do this, the tokens themselves take on reward value and can be used to establish and maintain any behavior of which the animal is capable. Animals, like men, can be trained to work for wages.

Once the animal is established in his "token economy," job training can proceed by means of a procedure known as "shaping," in which the animal is rewarded for increasingly accurate approximations of the desired response. In this way, animals can be taught to do things that are quite unlike anything they might do

³ With reference to his pigeon guidance system, Skinner (1960) wrote that "One virtue in crackpot ideas is that they breed rapidly and their progeny show extraordinary mutations."

spontaneously. Let us take the hypothetical case of a chimpanzee window washer.

Training would consist of the gradual establishment of the desired behavior by systematically rewarding the animal first when handling sponges and rags, then only when these implements were being applied to a model window, and finally only when they were being used in a manner that efficiently eliminated dirt from the window. When window-washing behavior had been well established, the animal would be taught to climb ladders in order to reach windows, to position ladders, and so on. Ultimately, the animal might be equipped with a device that would dispense a token whenever a window had been properly washed.⁴

Purely emotional objections notwithstanding, the major problem with which one is faced when contemplating implementation of ideas such as those presented here is that of what is "to be done" with the humans who are replaced by alternative organisms. A solution to this problem lies not in condemning the ideas and procedures that create it (progress usually results in disruption of the status quo), but in building the cultural and social resources and machinery needed to make it possible for all individuals to behave at a level commensurate with their intellectual endowment. This need not and should not result in an Orwellian civilization. Rather, the results of the judicious use of alternative organisms should allow for maximum individual freedom and development. We are advocating ideas that contain features designed to augment, not diminish,

the characteristics of man's existence, and it seems to us that this cause may be advanced when, as Cumming (1966) so beautifully put it, we learn when to "send a bird to do a bird's job [p. 247]."

REFERENCES

- BRELAND, K., & BRELAND, M. A field of applied animal psychology. *American Psychologist*, 1951, 6, 202-204.
- BRELAND, K., & BRELAND, M. The misbehavior of organisms. *American Psychologist*, 1961, 16, 681-684.
- CLARKE, A. C. Our dumb colleagues. *Harper's*, 1958, 216, 32-33.
- CUMMING, W. W. A bird's eye glimpse of men and machines. In R. Ulrich, T. Stachnik, & J. Mabry (Eds.), *The control of human behavior*. Glenview, Ill.: Scott, Foresman, 1966.
- DELGADO, J. M. R. Cerebral heterostimulation in a monkey colony. *Science*, 1963, 141, 161-163.
- DELGADO, J. M. R. Sequential behavior induced repeatedly by stimulation of the red nucleus in free monkeys. *Science*, 1965, 148, 1361-63.
- HERRNSTEIN, R. J. In defense of bird brains. *Atlantic*, 1965, 216, 101-104.
- HERRNSTEIN, R. J., & LOVELAND, D. H. Complex visual concept in the pigeon. In R. Ulrich, T. Stachnik, & J. Mabry (Eds.), *The control of human behavior*. Glenview, Ill.: Scott, Foresman, 1966.
- SKINNER, B. F. A case history in scientific method. *American Psychologist*, 1956, 11, 221-233.
- SKINNER, B. F. Pigeons in a pelican. *American Psychologist*, 1960, 15, 28-37.
- ULRICH, R., STACHNIK, T., & MABRY, J. (Eds.) *The control of human behavior*. Glenview, Ill.: Scott, Foresman, 1966.
- VERHAVE, T. The pigeon as a quality-control inspector. In R. Ulrich, T. Stachnik, & J. Mabry (Eds.), *The control of human behavior*. Glenview, Ill.: Scott, Foresman, 1966.
- WOLFE, J. B. Effectiveness of token-rewards for chimpanzees. *Comparative Psychology Monographs*, 1936, 12(Whole No. 60).

(Received January 30, 1969)

⁴ When training primates, shaping may not even be necessary. The systematic use of modeling procedures might be entirely sufficient for the establishment of many skills.

A SIMPLIFIED METHOD FOR INCREASING MAIL-SORTING EFFICIENCY

THOMAS REX LONG

Jacksonville University

DELL LEBO¹

*Child Guidance Clinic of Duval
County, Florida*

AND DOROTHY OGDEN

Jacksonville University

The number of pieces of mail sorted in a specified time was compared for two groups, each composed of 10 student volunteers. One group used standard post office department sorting techniques while the other used color-coded street names. Using a *t* test, the null hypothesis of no difference was rejected at the .05 level of confidence. The innovative technique was found to be superior to the conventional one.

Current mail-sorting techniques for mail carriers may not be as efficient as they might be. The sorting of mail is a slow and tedious process, especially for new carriers. They are faced with a sorting case with many street and box numbers on it, arranged according to the way they will be distributed along their route. Because of this arrangement, the case is confusing and the streets and addresses are difficult for the novice to find. Faced with this array and with several stacks of mail that must be sorted accurately, it is no wonder that initial sorting is slow and that speed is gained only after much practice.

Robert McFadden, a route carrier in Jacksonville, Florida, suggested that color coding a mail-sorting case could increase significantly the number of pieces of mail sorted in a given time. Such a procedure might also decrease the number of letters sorted erroneously. Hence, fewer letters would be delivered to incorrect addresses, which can result in a 1- or 2-day delay in delivery.

PROCEDURE

A standard mail-sorting case, such as that depicted in the *City Carrier's Instruction Handbook* (Bureau of Operations, no date), was used by both control and experimental groups. It was programmed for 12 streets and used seven shelves with forty 1-in. separations on each shelf. There were, thus, 280 slots for

sorting mail. The two groups were each composed of 10 student volunteers.

Prior to sorting, 500 envelopes were stuffed with dummy letters and addressed by hand by approximately 120 other students: Each student addressed 4 envelopes from the route to be used. This technique provided the sorter with a variety of handwriting such as a carrier might experience in actual practice. Care was taken to include at least one letter for each slot of the sorting case.

These letters were then distributed into six numbered stacks of mail for sorting. In order to eliminate bias with respect to any one stack of mail being favored as to the location of its letters in the sorting case, both groups sorted stacks in the same order. Thus, there was no opportunity for the letters themselves to offer a source of variation in the sorting process.

The control group sorted mail using the conventional technique as described in the *Handbook*. Each student was given a total of 10 consecutive trials consisting of 10 min. per trial. Two trials per day were given with the stipulation that a minimum of 2 hr. must intervene between trials. Trials were completed within 2 wk. Upon completion of each trial the average number of pieces of mail sorted and errors in sorting were noted.

The experimental group followed the procedure of the control group with major differences. These differences consisted of, first, using color-coded names for the 12 streets located on the sorting case. Since house numbers and streets were arranged by route in the case, colors appeared on various shelves of the case as well as, in some instances, running the length of the shelf. This change made the identification and location of the streets easier for someone new at the sorting task.

The sorting procedure was also changed for the experimental group in that they were asked to spend 10 min. memorizing or associating street names with their particular color before beginning sorting. The

¹ Requests for reprints should be sent to Dell Lebo, Child Guidance Clinic, 1635 St. Paul Avenue, Jacksonville, Florida 32202.

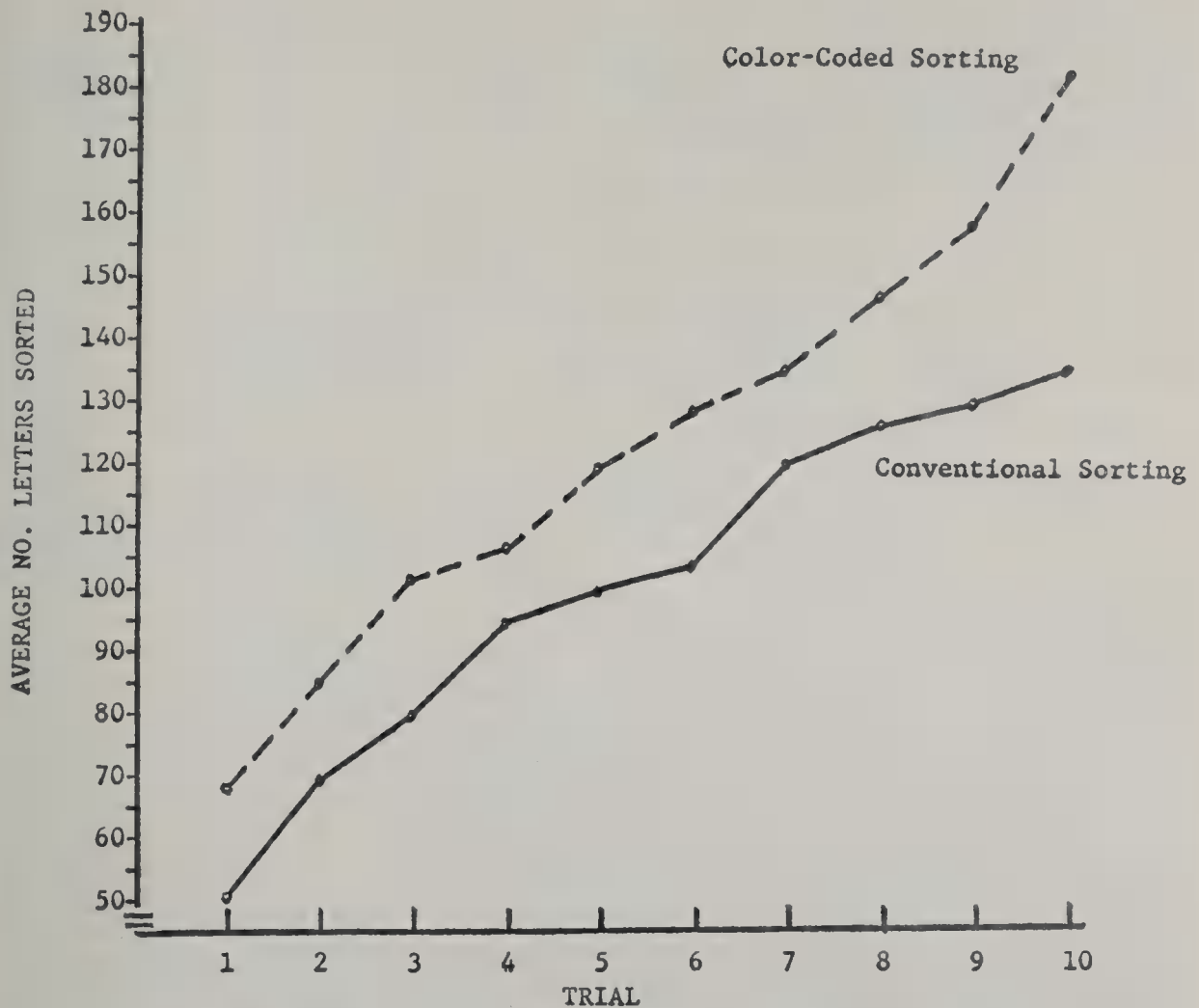


FIG. 1. Average number of letters cased or sorted per trial for those using the conventional post office method and those using the method of color coding the streets.

control group spent a similar 10-min. period simply reviewing the case as described in the *Handbook*. As the associations between colors and streets became stronger a new carrier should be able to sort mail faster because of a reduction in the time necessary to locate a street on the sorting case.

As incentives several rewards were offered the student volunteers. The first prize for both groups was a 3-day all-expense-paid trip to the Bahama Islands. A deep sea fishing trip was given to second and third place winners of both groups, and steak dinners were afforded "the losers."

RESULTS

Figure 1 clearly shows that the experimental group, who used color-coded street names, sorted more mail in each and every trial than did the control group using the conventional sorting method.

It is also apparent that the experimental group was learning the route at a faster rate than the control group. By the tenth trial the sorters using the color coding were sorting an average of 18.5 letters per minute, while those employing conventional sorting were sorting at a rate of 13.5 letters per minute. This difference was significant according to the t test at better than the .05 level of confidence.

Extrapolating these results to a 1-hr. period would suggest that those sorters using color-coded street names could sort 300 more pieces of mail than those employing the conventional sorting technique.

The experimental group also made fewer errors sorting mail. The difference was not statistically significant however.

The inference here is clear. New mail carriers could learn a sorting case more rapidly from the start and therefore sort more mail in a given period of time, if the street names on the sorting case were color-coded.

REFERENCE

BUREAU OF OPERATIONS, Distribution and Delivery Division. *City carrier's instruction handbook: M-41*. Washington, D. C.: United States Post Office Department, no date.

(Received January 27, 1969)

Manuscripts Accepted for Publication in the
Journal of Applied Psychology

- Individual and Situational Variables Relating to Job Satisfaction: W. W. Ronan*: School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332.
- Equity, Piece-Rate Overpayment and Job Security: Some Effects on Performance: Martin G. Evans* and Larry Molinari: School of Business, University of Toronto, 119 St. George Street, Toronto, 5, Canada.
- Demand Characteristics in a Field Experiment: Ned A. Rosen*: New York State School of Industrial and Labor Relations, Cornell University, Ithaca, New York 14850.
- Factor Analysis of Managerial Behavior Variables: J. C. Wofford*: School of Business Administration, University of Texas at Arlington, Arlington, Texas 76010.
- Relative Importance of Job Characteristics: W. W. Ronan*: School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332.
- Interpersonal Behavior in a Changing Organization: Daniel N. Braunstein*: College of Business Administration, University of Rochester, Dewey, Rochester, New York 14627.
- Feedback Analysis of Learning and Performance in Steering and Tracking Behavior: Karl U. Smith* and Vernon Putz: Department of Psychology, University of Wisconsin, Charter at Johnson, Madison, Wisconsin 53706.
- Relationship of Job Characteristics to Job Involvement, Satisfaction, and Intrinsic Motivation: Edward E. Lawler, III,* and Douglas T. Hall: Department of Administrative Sciences, Yale University, 2 Hillhouse Avenue, New Haven, Connecticut 06520.
- Need Satisfaction of Military Commanders and Staff: Vance F. Mitchell*: Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, British Columbia.
- Situational Influences on the Use of Power: Barry Goodstadt and David Kipnis*: Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.
- Laboratory Panel Estimation of Consumer Assessments of Taste and Flavor: William H. Bruvold*: School of Public Health, Earl Warren Hall, University of California, Berkeley, California 94720.
- Job Satisfaction Variables among Female Clerical Workers: Thomas R. Williamson*: D/459, Plt. H, Life Sciences Research Department, Goodyear Aerospace Corporation, 1210 Massillon Road, Akron, Ohio 44315.

* Asterisk indicates author for whom address is supplied.

A SITUATIONAL APPROACH TO LEADERSHIP EFFECTIVENESS

WALTER HILL¹

Department of Management, University of Florida

Fiedler's contingency model suggests that task-oriented leaders are more effective where the leadership situation is either very favorable or very unfavorable and that relations-oriented leaders are more effective in situations of intermediate favorability. This model was tested among supervisors in both interacting and coacting groups in two organizations. Results in the hypothesized direction were attained although they were not generally significant.

One of the most perplexing problems confronting managers has been to determine the leadership style most conducive to promoting effective work groups. Empirical studies directed toward finding that style which is most effective have yielded inconclusive and often contradictory results (Blake & Mouton, 1964; Fiedler, 1958; Lewin, Lippitt & White, 1939; Likert, 1961; Shaw, 1955). Although some theoreticians have been perplexed by the difficulty in identifying the one best leadership style, many practical supervisors have viewed the leadership literature with amusement as they have long recognized that both the directive, authoritarian, task-oriented leader and his counterpart, the democratic, human relations leader have proved effective in countless situations. The Contingency Theory of Leadership Effectiveness recently advanced by Fiedler (1964) suggests a theoretical explanation for both the confusion which now exists in the literature and the practical insights of many managers.

This theory suggests that leadership is an influence process where the ease or difficulty of exerting influence is a function of the favorableness of the group task situation for the leader. Although it has been recognized that the favorableness of each group task situation may depend on different variables, the three most commonly acknowledged determinants stated in their order of importance are leader-member relations, task structure, and position power. Once these variables have been measured, they can be ordered into

eight cells along a continuum to illustrate the relative degree of favorableness in a task situation as shown in Figure 1.

The most favorable situation exists when the leader enjoys good leader-member relations, is supervising a structured task, and possesses strong position power (Cell 1). The favorableness of the group task situation decreases as leader-member relations change from good to moderately poor; the most unfavorable situation is one where the leader-member relations are moderately poor, the task is unstructured, and position power is weak (Cell 8). The theory predicts that the task-oriented leader will be more effective in those situations which are either very favorable (Cells 1, 2, 3) or very unfavorable (Cell 8) and that the relations-oriented leader will be more effective in situations intermediate in favorableness (Cells 4, 5, 6, 7).

The empirical basis from which the contingency theory was induced is impressive: over 50 studies of 21 different types of groups. Recent studies (Blanchard, 1967; Fiedler, 1966; Hunt, 1967; Shaw & Blum, 1966) have tended to support this theory in interacting groups and have suggested that it may also be applicable in coacting groups (Hunt, 1967). The purpose of this study is to provide additional tests of the contingency model in interacting and coacting groups in real life organizations.

METHOD

An empirical test of the contingency model requires the following information: a measure of the supervisors' leadership styles, the classification of group supervisors into cells on the basis of leader-member relations, task structure and position power,

¹ Requests for reprints should be sent to the author, Department of Management, University of Florida, Gainesville, Florida.

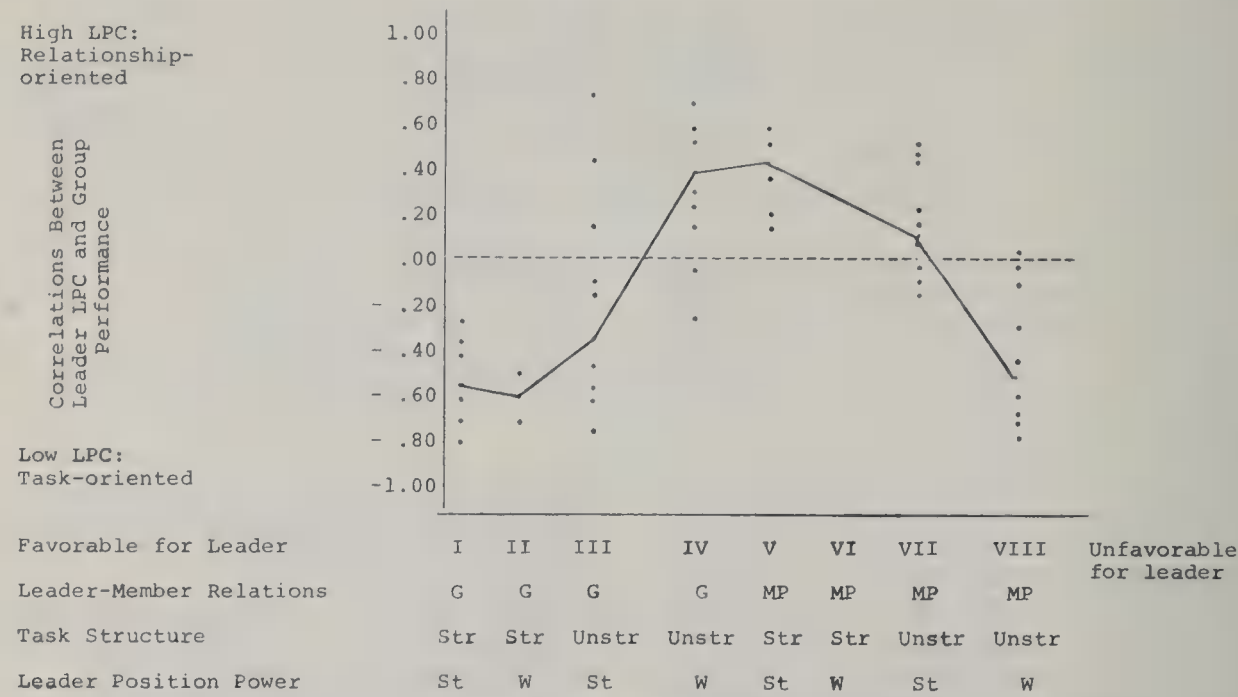


FIG. 1. Correlations between leaders' Least Preferred Co-worker scores and group effectiveness.

a measure of leadership effectiveness, and the determination of the correlations between leadership style and managerial effectiveness in each cell to be tested.

Supervisors were classified as task-oriented or relations-oriented on the basis of Fiedler's (1966) Least Preferred Co-worker (LPC) score. Leader-member relations were classified as good or moderately poor as a result of the leaders' perceptions of group atmosphere which were indicated by their responses to 10 semantic differential statements describing the group atmosphere (Fiedler, 1967). These responses were aggregated and divided into high and low group atmospheres (good and moderately poor leader-member relations) by taking the top and bottom third of the scores. Tasks were classified as structured or unstructured, and position power was defined as strong or weak by three judges completing questionnaires adapted by Hunt (1967). Leadership effectiveness was measured by asking Ss' immediate supervisors to rate their performance on relevant job duties and personal characteristics considered essential to job performance. Spearman's rank order correlation was employed to measure the relationship between leadership style scores and performance.

Subjects

The main consideration in the selection of organizations to be included in this study was that a sufficient number of supervised groups could be found performing both structured and unstructured tasks. This criterion was met in an electronics firm and a teaching hospital. The existing level of activity in the electronics firm enabled the author to investigate 28 groups performing unstructured state of the arts engineering tasks and 28 groups performing

structured assembly line operations. Since the nature of these tasks required a high degree of interdependency, the groups were judged to be interacting. The hospital afforded an opportunity to study 23 nursing groups whose tasks were judged to be unstructured and 25 groups performing structured tasks such as accounting, housekeeping, and routine maintenance. Since the hospital groups performed their work without a high degree of interdependency, they were judged to be coacting groups. The researcher met with the supervisors of these groups, explained that he was attempting to predict leadership effectiveness, and that the results of the study would be confidential. Twenty-eight assembly line foremen, 26 engineering supervisors, 23 nursing supervisors, and 25 managers from patient-support activities agreed to participate in the study and completed the required questionnaires.

RESULTS

Since the contingency model was tested in both interacting and coacting groups, results are reported and discussed separately.

Interacting Groups

Questionnaire returns from the electronics firm enabled the analysis of 28 structured and 26 unstructured task groups. Since supervisors of structured groups (assembly line foremen) were judged to have weak position power, a separation of these Ss into those having good and moderately poor leader-member relations allowed tests of the contingency

model in Cells 2 and 6. Since supervisors of unstructured groups (Engineering supervisors) were judged to have strong position power, a separation of these Ss into those having good and moderately poor leader-member relations enabled tests of the model in Cells 3 and 7. The calculation of Spearman's rho shown in Table 1 indicated correlations in the predicted direction for Cells 2, 3, and 7, although none of them reached an acceptable significance level. Cell 6 revealed a correlation in the opposite direction from that predicted by the model although it was not significant.

Coacting Groups

Questionnaire returns enabled the analysis of 25 structured and 23 unstructured task groups in the hospital. Since all supervisory positions were judged to have strong position power, the separation of managers of structured task groups (patient-supporting activities) into those having good and moderately poor leader-member relations allowed a test of the contingency model in Cells 1 and 5 while the division of unstructured task groups (nursing supervisors) enabled tests of the model in Cells 3 and 7. The calculation of Spearman's rho shown in Table 2 indicated that all correlations were in the hypothesized direction although only Cell 5 reached a significance level of .05.

DISCUSSION

Before the contingency model can be accepted as a valid theory of leadership effectiveness, many successful replications must be performed. Each cell in the model should be

TABLE 1

SPEARMAN RANK ORDER CORRELATIONS BETWEEN SUPERVISORS' LEAST PREFERRED CO-WORKER SCORES AND LEADERSHIP EFFECTIVENESS IN AN ELECTRONICS FIRM

| Cell tested | N ^a | Spearman's rho |
|-------------|----------------|----------------|
| 2 | 9 | -.097 |
| 3 | 8 | -.291 |
| 6 | 9 | -.238 |
| 7 | 8 | +.619 |

^a Since the number of structured and unstructured groups was not divisible equally into three groups, the remaining groups were assigned to the middle set and not included in the calculations.

TABLE 2

SPEARMAN RANK ORDER CORRELATIONS BETWEEN SUPERVISORS' LEAST PREFERRED CO-WORKER SCORES AND LEADERSHIP EFFECTIVENESS IN HOSPITAL STUDY

| Cell tested | N ^a | Spearman's rho |
|-------------|----------------|----------------|
| 1 | 7 | -.321 |
| 3 | 8 | -.214 |
| 5 | 7 | +.872* |
| 7 | 8 | +.529 |

^a Since the number of structured and unstructured groups was not divisible equally into three groups, the remaining groups were assigned to the middle set and not included in the calculations.

* $p < .05$.

treated as a separate hypothesis and all studies pertaining to a specific cell should be combined for purposes of ascertaining whether or not a correlation does exist. Only after it has been established that a correlation does exist will it prove fruitful to study the nature of the relationship through means of regression models. Thus, the studies reported in this paper can do no more than provide additional information concerning specific cells in the model.

Interacting Groups

The electronics firm investigation provided tests of the contingency model in Cells 2, 3, 6, and 7 in interacting groups as shown in Table 1. A comparison of the split-group correlations obtained in this study with the contingency model predictions indicates that the correlations in Cells 2, 3, and 7 support the hypothesis although none of the correlations reached a significance level of .05. The correlation obtained from Cell 6 was in the direction opposite to that predicted by the model.

Although none of the correlations in Cells 2, 3, and 7 reached an acceptable level of significance, they do fit generally into the results reported by Fiedler as shown in Figure 1, and thus a partial confirmation appears warranted. The discrepancies between the correlations reported in this work and those of Fiedler may be explained by several factors. First, the number of groups subjected to investigation in this study were smaller than those utilized by Fiedler. An acquaintance-

ship with statistical inference indicates that it is easier to reach a higher significance level if the sample size is greater. Thus, if more groups had been investigated, significant correlations might have been attained. Second, group effectiveness in Fiedler's original studies was always defined as measured performance stated in terms of such things as physical output, contest outcomes, and deviations from an intended target. The measures of group effectiveness employed in the studies reported in this article were based on effectiveness ratings by higher echelon superiors of the supervisors whose groups were studied. This procedure may introduce factors other than the actual performance of the group such as the bias of the evaluator. There was no way to measure this possibility. Thus, the method of effectiveness rating could account for some of the discrepancies between the correlations reported here and those discovered by Fiedler. Third, the definition of favorability used in the studies reported here was that originally espoused by Fiedler; that is, leader-member relations, task structure, and position power. Since the work performed was of a highly technical nature, it may be that the technical ability of the supervisor should have been a factor in the definition of the favorability dimension. The design of the study did not provide an opportunity to include this condition.

The negative correlation in Cell 6 can be explained much easier. An inspection of Fiedler's model in Figure 1 shows that no actual studies had ever been performed in this cell; the curve had merely been extended from Cell 5 to Cell 7. This extrapolation resulted in a prediction that high LPC leaders would be more effective than low LPC leaders where moderately poor leader-member relations existed, the task was structured, and position power was weak.

Since this study is the first to measure such conditions, it may be that the extrapolation of the model was unwarranted. There obviously is no reason why the curve cannot dip below the line in Cell 6 and rise again in Cell 7. In fact, it may be argued that the existence of moderately poor leader-member relations raises the anxiety level of the high LPC leader since good relations are of primary

importance to him and his reactions intended to develop better relationships actually interfered with the performance of a task which was basically well structured and required better direction. On the other hand, a low LPC leader is not as alarmed by the existence of moderately poor leader-member relations and continues focusing his attention on task performance improvements in the structure which may lead to greater effectiveness.

It must be remembered that this model suggests only that one type of leader tends to be more effective than another type in a given cell on the favorability continuum. This implies that special conditions may enable the latter type of supervisor also to be effective. Thus, the model only suggests that, *ceteris paribus*, one type is more likely to succeed than another.

Coacting Groups

The study conducted in the hospital provided tests of the contingency model in Cells 1, 3, 5, and 7, in coacting groups. The results from the hospital study indicate that all correlations are in the predicted direction although a significance level of .05 was attained only in Cell 5 as shown in Table 2. This suggests that when leader-member relations are moderately poor, the task is unstructured, and position power is strong that high LPC leaders are more effective than low LPC leaders. This conclusion is consistent with the literature which generally holds that a supervisor dealing with professional people who perform unstructured tasks should adopt a democratic rather than an authoritarian leadership style.

Although correlations in cells other than Cell 5 did not reach an acceptable level of significance, it may be that the reasons advanced previously can account for the lack of significance. Certainly, the model has intuitive appeal and warrants further replication.

REFERENCES

- BLAKE, R. B., & MOUTON, J. S. *The Managerial Grid*. Houston: Gulf Publishing Company, 1964.
- BLANCHARD, K. H. College boards of trustees: A need for directive leadership. *Academy of Management Journal*, 1967, 10, 409-417.
- FIEDLER, F. E. *Leader attitudes and group effectiveness*. Urbana: University of Illinois Press, 1958.

- FIEDLER, F. E. A contingency model of leadership effectiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 1, New York: Academic Press, 1964.
- FIEDLER, F. E. The effect of leadership and cultural heterogeneity on group performance: A test of the contingency model. *Journal of Experimental Social Psychology*, 1966, 2, 237-264.
- FIEDLER, F. E. A theory of leadership effectiveness. New York: McGraw-Hill, 1967.
- HUNT, J. G. A test of the contingency model in three organizations: Urbana: University of Illinois Press, 1967.
- LIKERT, R. New patterns of management. McGraw-Hill, 1961.
- LEWIN, K., LIPPITT, R., & WHITE, R. K. Patterns of aggressive behavior in experimentally created "social climates." *Journal of Social Psychology*, 1939, 10, 271-299.
- SHAW, M. A comparison of two types of leadership in various communication nets. *Journal of Abnormal and Social Psychology*, 1955, 50, 127-134.
- SHAW, M., & BLUM, J. M. Effects of leadership style upon group performance as a function of task structure. *Journal of Personality and Social Psychology*, 1966, 3, 233-242.

(Received February 3, 1969)

RELEVANCE OF RATER-RATEE ACQUAINTANCE IN THE VALIDITY AND RELIABILITY OF RATINGS¹

NORMAN E. FREEBERG²

Educational Testing Service

Unacquainted Ss worked in three-man groups under relevant (mathematical tasks) and irrelevant (socializing) acquaintance conditions. The Ss rated one another on scales that defined several cognitive skills. They were also rated on these same scales by observers who were dependent on visual information, exclusively, and were unacquainted with the group members or the specific nature of the tasks being performed. As hypothesized, group members under the relevant acquaintance condition achieved consistently good validity for all three cognitive areas—with the best validity for ratings of math ability. Validity under the irrelevant acquaintance condition was nil on all scales. Observers, surprisingly, achieved significant validity (although at lower levels than participating group members) for ratings under the relevant acquaintance condition. Levels of inter- and intrarater reliability were not associated with levels of validity under the various rating conditions.

Efforts devoted to improvement of the rating process have focused primarily upon content, construction, and format of the rating instrument. Only limited attention has been paid to the nature of the rater-ratee relationship—in terms of the conditions under which raters make their observations or the availability of behavioral cues relevant to the characteristics being judged.

The importance of this aspect of rating has not gone unnoticed in general discussions by such writers as Thorndike and Hagen (1961) who conclude that “the ideal rater is the person who has had a great deal of opportunity to observe the person being rated in situations in which he would be likely to show the qualities on which ratings are desired [Ch. 13].” Burt (1942) has similarly stressed “the conditions under which the ratee has been observed” as affecting the accuracy of rating and suggests that, in addition to requiring information regarding how long or how well the rater knows the ratee, one might also request information regarding the circumstances under which he was known. In his *mathematico-deductive Theory of Rating*, Wherry (1952) derives a theorem which specifies that “raters will vary in the accuracy

of ratings given in direct proportion to the relevancy of their previous contact with the ratee [p. 10].”

Empirical verification of these logical contentions is relatively scarce. Several studies have dealt with a concept of rater-ratee acquaintance based largely upon the length of time that the ratee was known by the rater (Ferguson, 1949; Knight, 1923; United States Department of the Army, 1952). Findings from these studies indicate that with an increasing degree of acquaintance there tends to be an increased “leniency” (i.e., skewness of trait distributions toward the favorable end of the scale) and higher intercorrelations among traits which is assumed to be the product of greater halo effect. Madden (1961) discusses a similar effect for raters evaluating jobs with which they have varying degrees of familiarity. Increased rater reliability is also shown with increases in “opportunity to observe” (United States Department of the Army, 1952) or greater degree of acquaintance (Ferguson, 1949; Kornhauser, 1926). However, failure to find such significant increases in reliability with increasing acquaintance is reported by Mays (1954) and Hollander (1956, 1957) for peer ratings obtained from OCS classes. Rater reliability in the form of interrater consistency in one such study generally remained high throughout a 13-wk. period as did agreement between early and later peer nominations (Hollander, 1957).

¹ The research reported in this paper is based, in part, on data utilized for a doctoral dissertation submitted to the Ohio State University.

² Requests for reprints should be sent to the author, Developmental Research Division, Educational Testing Service, Princeton, New Jersey 08540.

Claims for improved validity and better trait discrimination with increased familiarity are reported for supervisor ratings of subordinates (Ferguson, 1949) and students' ratings of teachers (Bare, 1954). But again, there are dissident findings that serve to hint at the complexity of the problem and the probable effects of rating conditions. An early study by Moore (1937) indicated no improvement in teachers' ranking of their students' intelligence over one to three years of acquaintance; while Hollander (1956) points out that, although peer nominations did show increased average validity with increasing length of acquaintance, the validity is stable over time if degree of friendship is partialled out of the nominations. In any of the above studies there is little or no control over the specific degree and character of the rater-ratee relationship as well as serious limitations in the scope or objectivity of available criteria used to demonstrate validity.

One solution to such inadequacies is to organize individuals into groups, wherein reasonable control over the extent and nature of rater-ratee contact can be exercised. When this is done, trait behaviors to be evaluated can also be elicited, in terms of those specific behaviors that define any given trait, and for which objective criteria are readily available. By means of such an approach, the present study attempts to deal with the relevance of rater-ratee contact as it affects the reliability and validity of rater judgments.

METHOD

Subject Groups

Participants. Sixty-nine undergraduate males were organized into 23 groups, with each group composed of three members referred to as "participants." The first experimental session in which each group served was designated as the Relevant Contact condition. From the same pool of 69 Ss, 21 three-man groups were reorganized for a second session designated as the Irrelevant Contact condition (wherein none of the three participants had previously served with one another under the Relevant Contact condition). Participants in any given group were totally unacquainted with one another prior to the group sessions. All identification during the group sessions, for rating or any other purpose, was by reference to different colored laboratory coats worn by each participant.

Observers. For both experimental sessions, two male students designated as "observers" viewed the ac-

tivities of the three-man participant groups through a one-way vision screen. A different pair of observers viewed every session; they were completely unacquainted with the participants they observed.

Experimental Conditions

Relevant contact. The Ss in the participant group entered a room containing a separate table and chair to be used by each man and a large drawing board near the center of the room to be used by the three men as a group. The first 10 min. were spent by each S working independently at one of the tables on a set of arithmetic and algebraic problems under instruction to complete as many of the problems as possible. This was followed by an additional 10-min. session during which the three Ss were to compare their answers and discuss any discrepancies in solutions. They were then given a set of similar problems at a more difficult level and instructed to solve them, as a group, while working together at the drawing board. The Ss were to arrive at a single answer for each problem and to complete as many problems as possible during an allotted 30 min.

The rationale for individual problem solving and comparison of solutions, prior to working the problems as a group, was to expose Ss to one another's arithmetic ability. This was intended to maximize the opportunity for each S to display "relevant" (i.e., arithmetic) skills and minimize dominance of group activity by an aggressive, but less capable, group member. The requirement for one agreed-upon solution was also intended to serve this purpose.

Included in the instructions to Ss was the fact that they would be in competition with other groups and that monetary prizes would be awarded to members of groups which solved the greatest number of problems correctly. It was explained that the "mirrors" at each end of the room were one-way observation screens and that they would be observed during the course of the session.

Observers behind the one-way screens were restricted to visual cues *only* and were unable to hear any conversation in the room. From their viewing position, they could not observe what the participants were writing either at the tables or at the drawing board. Thus the specific nature of the activity should have been relatively unknown to them and, in fact, these observers were intended to represent raters having minimal contact or relevant cues.

Irrelevant contact. At least two weeks from the time of the Relevant Contact sessions, three-man groups, reorganized from the original pool of Ss, returned to participate in a second session of group activity designated as the Irrelevant Contact condition (i.e., irrelevant to arithmetic performance). The task chosen required that the three Ss, working as a group, construct an "artistic" product of some sort from the material in Tinker Toy sets. The intent in these sessions was to provide a task that would elicit behavioral cues largely irrelevant to the traits under evaluation. As during the relevant contact sessions, observer raters were given no information other than what they could derive from visual observation of the group's activity in the test room.

ability), and academic grades in the form of the cumulative point-hour ratio.

Data Analysis

From each of the experimental sessions, a pair of rating scores was obtained for each participant (ratee) on each of the three scales (Math, Academic, and Intellectual scales). These pairs of ratings, which were obtained from the two observers and the ratee's two fellow participants, were summed to provide a single score. Thus, for each ratee there was a total of 12 rating scores derived from the combination of three scales, used by two types of raters, under two contact conditions. Along with the three criterion measures, these variables were intercorrelated in a 15×15 matrix factor that was analyzed by the Thurstone (single) Group Centroid Method (Thurstone, 1947). Since there were pairs of rating scores for each scale under each rating condition, the degree of agreement between raters (interrater reliability) could be determined by intraclass correlations. For measures of rater self-consistency (in essence, scale reliability) Kuder-Richardson (21) estimates were obtained. Thus, these two forms of rater reliability could be contrasted with the degree of rater validity achieved for the same rating scales under the same rating conditions.

| | P | | | | | | O | | | | | | OSPE 13 | Point hour 14 | Math. test 15 |
|----|-------------------|--------|--------|--------------------|--------|--------|------------------|--------|--------|--------------------|---------|---------|------------|---------------------|---------------------|
| | Relevant contact | | | Irrelevant contact | | | Relevant contact | | | Irrelevant contact | | | | | |
| | M 1 | A 2 | I 3 | M 4 | A 5 | I 6 | M 7 | A 8 | I 9 | M 10 | A 11 | I 12 | | | |
| | Intercorrelations | | | | | | | | | | | | | | |
| 1 | | 72 | 76 | 01 | 17 | 14 | 56 | 49 | 57 | 08 | 11 | 19 | 55 | 44 | 60 |
| 2 | -02 | | 77 | --03 | 11 | 08 | 49 | 50 | 56 | 09 | 11 | 21 | 42 | 43 | 41 |
| 3 | 01 | 04 | | -04 | 12 | 07 | 33 | 35 | 41 | 12 | 15 | 27 | 51 | 46 | 42 |
| 4 | 00 | 01 | -01 | | 71 | 73 | 04 | 05 | 12 | -02 | 04 | 05 | 05 | 01 | -05 |
| 5 | -03 | 00 | 02 | 00 | | 70 | 19 | 11 | 16 | 04 | 22 | 10 | 08 | 04 | 16 |
| 6 | -01 | -01 | -03 | 00 | 00 | | 05 | 00 | 13 | 00 | 08 | 01 | 09 | 01 | 09 |
| 7 | 04 | 02 | -02 | -02 | 02 | -02 | | 72 | 81 | 07 | 14 | 00 | 32 | 26 | 43 |
| 8 | 02 | 02 | -03 | 03 | 01 | -02 | 01 | | 82 | 13 | 19 | 16 | 28 | 23 | 27 |
| 9 | 01 | 02 | -02 | -01 | -04 | 02 | 01 | 00 | | 03 | 11 | 02 | 30 | 31 | 38 |
| 10 | 00 | 00 | 01 | 00 | -05 | 00 | 00 | -02 | 01 | | 63 | 76 | 18 | 04 | -02 |
| 11 | -02 | 00 | 03 | 05 | 04 | -01 | -01 | 01 | -01 | 00 | | 68 | 26 | 18 | 09 |
| 12 | 02 | 02 | -01 | 02 | -01 | -04 | -03 | 04 | 00 | 01 | 02 | | 32 | 25 | -02 |
| 13 | -02 | -02 | 01 | 02 | 05 | -01 | 00 | 03 | -02 | 02 | -01 | 00 | | 65 | 44 |
| 14 | -04 | 03 | -02 | 02 | 00 | -04 | -02 | 00 | 00 | -01 | -02 | 02 | 01 | | 40 |
| 15 | 02 | -30 | 01 | -02 | 01 | 00 | 02 | -03 | 01 | -01 | 02 | -02 | 01 | 01 | |
| | Residuals | | | | | | | | | | | | | | |

TABLE 2

ROTATED ORTHOGONAL FACTOR LOADINGS AND COMMUNALITY ESTIMATES

| Variable | Factors | | | | | | |
|------------------------------------|---------|-----|-----|-----|-----|-----|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 | h^2 |
| 1. Scale M—Relevant Contact (P) | 61 | 61 | 29 | 02 | −02 | −03 | 83 |
| 2. Scale A—Relevant Contact (P) | 49 | 51 | 45 | −04 | 07 | 02 | 71 |
| 3. Scale I—Relevant Contact (P) | 58 | 37 | 58 | −05 | −05 | 05 | 81 |
| 4. Scale M—Irrelevant Contact (P) | 13 | −14 | 00 | 86 | 08 | −03 | 74 |
| 5. Scale A—Irrelevant Contact (P) | 21 | 12 | −06 | 81 | −01 | 06 | 71 |
| 6. Scale I—Irrelevant Contact (P) | 20 | 00 | 05 | 82 | −07 | −04 | 71 |
| 7. Scale M—Relevant Contact (O) | 36 | 53 | 03 | 04 | 56 | −02 | 72 |
| 8. Scale A—Relevant Contact (O) | 27 | 48 | 13 | 01 | 65 | 10 | 75 |
| 9. Scale I—Relevant Contact (O) | 38 | 49 | 13 | 10 | 72 | −06 | 93 |
| 10. Scale M—Irrelevant Contact (O) | 09 | 10 | −04 | 01 | −01 | 85 | 74 |
| 11. Scale A—Irrelevant Contact (O) | 27 | 04 | −13 | 09 | 06 | 70 | 56 |
| 12. Scale I—Irrelevant Contact (O) | 30 | −05 | 14 | 02 | −06 | 86 | 68 |
| 13. O.S.P.E. | 80 | 09 | 00 | −07 | −02 | 09 | 65 |
| 14. Point hour | 80 | −03 | 04 | −14 | 04 | −01 | 62 |
| 15. Math. test | 50 | 48 | −09 | −02 | −06 | −13 | 45 |

RESULTS AND DISCUSSION

Rater Validity

The intercorrelations presented in Table 1 show a pattern of rater validity that is obviously superior for the condition under which observable behaviors are relevant to the characteristics being rated. The highest rater validity under the Relevant Contact condition ($r = .60$) occurs for participants rating one another on mathematical ability. In fact, all of the scales used by participants under the Relevant Contact condition possess significant validity for all three criteria. Logically, this is a result of halo effect and reflects the interrelation among the three criterion measures of academic, intellectual, and mathematical abilities. Thus, a rater who must depend, primarily, upon his observations of one area of a ratee's abilities (i.e., mathematics skill) uses such available behaviors as the basis for judging other cognitive skills. His accuracy (validity) in making judgments of these other *nondirectly* observed abilities is, apparently, dependent upon the extent to which observed and nonobserved abilities are interrelated.

Observers, under the Relevant Contact condition, also achieved significant validity with each scale but at a lesser level of validity than the participants (r 's ranging from .23 to .38).

Given the rather minimal information that they possessed, the achievement of any level of accuracy is surprising and of particular interest for subsequent discussion.

It is, however, the factor structure for all ratings under the various rating conditions that provides the clearest picture of the overall effects. Six factors were extracted, based on a criterion of reducing residual values to approximately zero (the residual matrix of Table 1 indicates that the criterion was, essentially, satisfied). Factors were rotated, manually, to psychological meaningfulness. The first factor could be classed as a general one representing accuracy in judging *General Cognitive Ability*, since the highest loadings appear for the three criterion measures and most of the other variables achieve positive loadings on this factor. Participants under the Relevant Contact condition display the greatest capability to make such judgments. The next highest contribution to the factor appears for the ratings by observers under the Relevant Contact condition, while ratings under the Irrelevant Contract condition show the least contribution to this dimension.

Factor 2 provides the clearest pattern of the effectiveness of a relevant acquaintance condition on rater validity. The factor is best designated as *Computational (Arithmetic) Ability*, with the Mathematics Test as the

only criterion measure loading on the factor and loadings of interpretable magnitude for rating conditions occurring only under the Relevant Contact condition (participants and observers). Ratings made on the basis of ratee performance irrelevant to the characteristics being rated make virtually no contribution to this factor. Essentially, then, raters who have had an opportunity to deal with ratees in some form of arithmetic skills are unquestionably more accurate in making judgments of such skills.

The four remaining dimensions are striking examples of "specific halo," that is, specific to each of the four rating conditions. Thus, Factor 3 defines halo effect for Relevant Contact participants; Factor 4 defines it for Irrelevant Contact participants; Factor 5 is halo specific to observer ratings under the Relevant Contact condition; and Factor 6 represents halo for observers under the Irrelevant Contact condition. From the magnitude of the loadings it can be seen that variance attributable to specific halo is largest under the Irrelevant Contact conditions.

This illustration of a strong halo effect for the three rating scales under each rating condition would lend further support to the above contention that halo served to spread validity, under the Relevant Contact conditions, from observation of arithmetic performance to ratings of related academic and intellectual performance. Halo effect may, therefore, be said to lend validity to nonobserved characteristics insofar as these are related areas of performance not directly observed (for example, one may display competence in rating an individual's athletic ability for numerous sports although having observed his performance in only one sport). Bingham (1939) has previously drawn attention to this positive role of rater generalization in terms of wanted and unwanted halo rather than the customary conception of all halo as indiscriminated "blur."

But the question remains of how observers were able to achieve a significant degree of rating accuracy, under the Relevant Contact condition, despite the minimal information they possessed and on the basis of which it was assumed their rating validity should be nil (i.e., they could be considered as somewhat of a control group of raters). Observer

visual information was almost entirely dependent upon fairly gross, group-interactive behavior such as conversation directed by one member to others, explanatory gestures, an individual's position at the drawing board demonstrating something, writing, passively serving, etc. (Lip reading by observers is assumed to have provided no significant information.) These are forms of group behavior which allow for reasonably good definition of the degree of dominance exhibited by each member of the participant group. As indicated in the Method section, specific efforts were made to enhance group-member awareness of which individual possessed the greatest arithmetic competence. The most logical assumption, therefore, is that this degree of observed dominance served as the primary dimension along which ratings of arithmetic, academic, or intellectual ability were made by observer raters.

If group-member dominance does tend to be associated with a higher level of ability for the tasks involved, as has been suggested by Shevitz (1955), then the explanation of these results is tenable.³

Rater Reliability

Coefficients for intra- and interrater reliability are shown in Tables 3A and 3B for each scale, under each rating condition.

Internal consistency of the scales—which can be considered as a form of intrarater reliability—is fairly good over all conditions, with KR-21 estimates ranging from .56 for participants rating one another on academic ability under the Irrelevant Contact condition to .79 for observer ratings of participants' math ability under the Irrelevant Contact condition. Contrasting these reliability coefficients with validities of Table 1 indicates that intrarater reliability in a given rating situation obviously has little bearing on rater validity achieved.

Agreement between raters (interrater reliability) is found, overall, to be at a more

³ For each experimental session, *E* spent almost the entire time with the two observer raters and would conclude that, other than the physical characteristics of the participants, group interactive behaviors were virtually the only visual cues that could be utilized reasonably by an observer asked to make such judgments of cognitive skills.

TABLE 3
INTRA- AND INTERRATER RELIABILITY

| Condition | Scale | | | <i>N</i> |
|---|------------------|------------------|------------------|----------|
| | Math | Academic | Intelligence | |
| Intrarater Reliability (Kuder-Richardson, 21) | | | | |
| Relevant Contact (P) | .71 | .57 | .64 | 69 |
| Relevant Contact (O) | .74 | .71 | .61 | 69 |
| Irrelevant Contact (P) | .68 | .56 | .66 | 63 |
| Irrelevant Contact (O) | .79 | .77 | .66 | 48 |
| Interrater Reliability (Intraclass Correlations) | | | | |
| Relevant Contact (P) | .52 | .33 | .34 | 69 |
| Relevant Contact (O) | .38 | .32 | .34 | 69 |
| Irrelevant Contact (P) | .23 ^a | .06 ^a | .15 ^a | 63 |
| Irrelevant Contact (O) | .46 | .60 | .46 | 48 |

^a Correlation not significantly greater than zero.

modest level with r 's ranging from .06 to .60. Again, one could hardly make a case for a pattern of interrater agreement that coincides with rater validity, particularly since the highest interrater reliabilities appear for the three scales used by observers under the Irrelevant Contact condition, where information available to raters was minimal, as was their resulting validity. In effect, the evidence tends to support Wherry's (1952) contention that "the reliability of a rating scale tells us very little about its value, since the apparent reliability may be due to bias rather than true score [p. 39]."

CONCLUSIONS

Where the nature of rater-ratee contact is reasonably controlled, a situation which provides an opportunity for more relevant cues to be exhibited has been shown to result in more accurate ratings. Rater reliability, however, in the form of either rater self-consistency or agreement between raters is not necessarily affected by the relevance of the observations to the qualities judged.

Raters, as might be expected, utilize observed cues relevant to one area of ability in order to generalize to other nondirectly observed characteristics. The resultant effect of

such halo can be one of enhancing validity for ratings of the nonobserved characteristics, insofar as these are related to the ones observed. Where ratee behaviors elicited are logically irrelevant to the characteristics being judged, the ratings have been shown to lack validity over several related criterion characteristics. Raters thus generalize their inaccuracies in rating (based upon irrelevant cues), as well as their judgments based upon relevant observations.

In practice, then, it seems that ratings dependent upon mere acquaintance of rater and ratee are of doubtful value without specific knowledge of the nature of this acquaintance. If length of rater-ratee acquaintance is used to justify the choice of raters, this can only prove defensible where longer acquaintance provides greater opportunity to observe behaviors that are relevant to the traits judged. Nevertheless, many critical judgments of such qualities as an individual's "patriotism" or "reliability" as an employee, are made by raters whose contact is often no more relevant than that of a neighbor or a social acquaintance. Further, agreement between raters in such practical rating situations is often used to bolster the claim of rating accuracy. Here again, as has been pointed out, such agree-

ment between individuals (or self-consistency by individual raters) is not at all a necessary reflection of rater validity.

In addition to such general conclusions drawn from these study data, it is felt that an important consideration is the value of the method used. Achieving reasonable control over the nature and extent of rater-ratee contact, by use of selected individuals in a small group setting, allows for variation of a number of parameters that might be influenced by the conditions of acquaintance. A next research step would, logically, be a more precise delineation of degree of relevance as it affects rating accuracy over a greater variety of traits, from specific cognitive skills to aspects of social or attitudinal characteristics.

REFERENCES

BARE, R. H. Bias as related to rater contacts. Unpublished doctoral dissertation. Ohio State University, Columbus, Ohio, 1954.

BINGHAM, W. V. Halo, invalid and valid. *Journal of Applied Psychology*, 1939, 23, 221-228.

BURTT, H. E. *Principles of employment psychology*. New York: Harper, 1942.

FERGUSON, L. W. The value of acquaintance ratings in criterion research. *Personnel Psychology*, 1949, 2, 93-102.

HOLLANDER, E. P. The friendship factor in peer nominations. *Personnel Psychology*, 1956, 9, 435-447

HOLLANDER, E. P. The reliability of peer nominations

under various conditions of administration. *Journal of Applied Psychology*, 1957, 41, 85-90.

KNIGHT, F. B. The effect of the "acquaintance factor" upon personnel judgments. *Journal of Educational Psychology*, 1923, 14, 129-142.

KORNHAUSER, A. W. Reliability of average ratings. *Journal of Personnel Research*, 1926, 5, 309-317.

MADDEN, J. M. A further note on the familiarity effect in job evaluation. USAF, Lackland Air Force Base, Texas, ASD-TN-61-47, June, 1961.

MAYS, R. J. Relationships between length of acquaintance and nature of trait rated and agreement between raters. United States Air Force Personnel Training and Research Center, Research Bulletin No. 54-55, 1954.

MOORE, C. C. Note on the accuracy of judgment in pupil ability ratings. *Journal of Educational Psychology*, 1937, 28, 391-393.

SHEVITZ, R. N. An investigation of the effects of exclusive possession of information upon attempts to lead in small groups. Unpublished doctoral dissertation. Ohio State University, Columbus, Ohio, 1955.

THORNDIKE, R. L., & HAGEN, E. *Measurement and evaluation in psychology and education*. (2nd ed.) New York: Wiley, 1961.

THURSTONE, L. L. *Multiple-factor analysis*. Chicago: University of Chicago Press, 1947.

United States Department of the Army. The effect of opportunity to observe on the reliability and independence of ratings. Personnel Research Branch, Research Note No. 14, 1952.

WHERRY, R. J. The control of bias in rating: VII. *A theory of rating*. Department of the Army, Personnel Research Branch, Report No. 922, Final Report, February 1952.

(Received February 3, 1969)

EFFECTS OF ORGANIZATIONAL STRUCTURE ON CORRELATIONS BETWEEN MEMBER ABILITIES AND GROUP PRODUCTIVITY¹

GORDON E. O'BRIEN²

University of Illinois

AND

A. G. OWENS

*Psychology Research Unit,
Australian Military Forces*

In two different studies it was found that the contribution of member ability to group productivity was dependent on both the ability of the member and the kind of task organization employed by the group. The first study was carried out in a military setting with 40 four-man groups and the second study involved 48 three-man groups with undergraduate college students as Ss. When the group task required members to cooperate by coordinating their efforts, group productivity was significantly affected by both the average ability of the group and the ability of the dullest member. When the group task required members to cooperate by collaborating, group productivity was not significantly affected by either the average ability of the group or the ability of the dullest member.

A number of reviews (Gibb, 1954; Heslin, 1964; Mann, 1959) have shown that the abilities of group members are generally related to group productivity in a positive manner. Correlations between measures of task-relevant abilities and group productivity are typically small, however. One possible reason for the smallness of these correlations is the neglect by researchers of the organization used by members in performing the group task. The organization most used in studies of small group performance is a collaborative one where group members are expected to cooperate with each other at all stages of the task activity (e.g., discussion and problem-solving task). Under these conditions, it has been found by some researchers that personality factors are better predictors of group

productivity than task-relevant abilities (Schutz, 1958). Members with superior ability are often unable to contribute significantly because of personal conflicts and incompatibility with other members.

Little is known about the relationship between abilities and group productivity in situations where the group is required to cooperate through task coordination rather than through collaboration. Coordination occurs when different tasks are allocated to different positions and the tasks are then ordered by definite precedence relationships. Under these conditions all members not only have an opportunity to influence the group product, but are actually required to contribute. Hence, if members of a group are allocated separate tasks of equal importance, it is likely that the group product will be proportionate to their summed abilities. Furthermore, because of the definite task sequencing, it is probable that the quality of the group product would be particularly sensitive to poor performance by any one person. This form of cooperation is observed in assembly lines where shoddy performance by one worker often results in an inferior product, even though the remaining members are quite competent.

In summary, for tasks where coordination is high, group productivity should be positively related to the summed abilities of all

¹The study was supported in part by Contract R 177-472 with the Advanced Research Projects Agency (Fred E. Fiedler and Harry C. Triandis, principal investigators). Alice J. Sanders and Anthony Biglan assisted with the data analysis.

The authors acknowledge the assistance of the Australian Military Board in allowing the study reported to be carried out at Puckapunyal military camp, Victoria. The opinions expressed are those of the authors and do not reflect any view official or unofficial of the Australian Department of Army.

²Requests for reprints should be sent to Gordon E. O'Brien, who is now at the Psychology Department, School of Social Sciences, Flinders University, Bedford Park, South Australia 5042.

group members and positively related to the ability of the least competent member. For tasks with high collaboration, group productivity should be less strongly related to these ability measures. Evidence to support these statements was obtained from two experiments which were concerned with the relationships between group structure and productivity. As measures of member ability were available, the effects of these abilities upon group output could be estimated.

METHOD

Coordination and Collaboration

In order to measure the amount of coordination or collaboration required by a given organization structure, two cooperation indices were derived (O'Brien, 1968; Oeser & O'Brien, 1967; Witz &

O'Brien, in press). These indices can be used whenever it is possible to identify the positions and tasks in a group, the allocation relationships ordering tasks and positions, and the precedence relationships ordering tasks.

The collaboration index, C_L , is given by the formula

[1]
$$C_L = \frac{\sum_{i=1}^{i=m} \sum_{j=1}^{j=n} (p_{ij}t_j) - n}{n(m-1)}$$

where $\sum_{i=1}^{i=m} \sum_{j=1}^{j=n} (p_{ij}t_j)$ is the sum of the entries in the

task allocation matrix (PT). The (ij) entry in this matrix has the value 1 if Position p_i has allocated to it Task t_j and the value 0 if t_j is not allocated to p_i ; n = the number of subtasks and m = the number of positions.

An index of strict coordination, C_o , is given by the formula

[2]
$$C_o = \frac{\sum_{i=1}^{i=m} \sum_{j=1}^{j=m} (x_{ij}y_j) - \sum_{i=1}^{i=m} \sum_{i=1}^{i=m} (x_{ii}y_i) + \sum_{i=1}^{i=m} \sum_{j=1}^{j=m} (u_{ij}v_j)}{M(m)M(n)}$$

where the entries of $(x_{ij}y_j)$ and $(x_{ii}y_i)$ are obtained from the resultant of the following matrices

$$(PT) \cdot (TT) \cdot (PT)' - (PT) ((TT) \cdot (PT)' \circ (PT)') - ((PT) \cdot (TT) \circ (PT)) \cdot (PT)'$$

$(PT)'$ is the transpose of (PT) and the symbol \circ indicates elementwise multiplication. The precedence matrix is (TT) . The (ij) entry in this matrix has the value 1 if Task t_j must be preceded by Task t_i and the value 0 if Task t_j is not preceded by Task t_i . The entries of $(u_{ij}v_j)$ are obtained from the resultant of the following matrices

$$((PT)' \cdot (PT) \circ (PT)' \cdot (PT)) \circ (TT) - ((PT) \circ (PT))' \cdot ((PT) \circ (PT)) \circ (TT).$$

$$M(m) = \frac{1}{4} m^2 \text{ when } m \text{ is even and } M(m) = \frac{1}{4} (m^2 - 1) \text{ when } m \text{ is odd.}$$
$$M(n) = \frac{1}{4} n^2 \text{ when } n \text{ is even and } M(n) = \frac{1}{4} (n^2 - 1) \text{ when } n \text{ is odd.}$$

The indices were used to calculate the collaboration and coordination values for the task organizations used by groups in the following studies.

Study I—Army Study

In this study 160 Australian regular army soldiers (NCOs and privates) were assigned to 40 four-man groups. Twenty of the groups were given the task of writing a recruiting letter and the remaining groups were required to prepare two charts showing the results of apprentice examinations at Army technical schools. Groups in each set of 20 were matched on status or rank structure and prior

acquaintance. For each group, the leader was defined as the soldier with highest rank.

Recruiting letter task. For this task, the group was asked to write a letter to Australians in the age group 17–20 yr. old. Group members were told that the letter should explain why the Army is a worthwhile career and should encourage them to enlist in the Australian Regular Army. Instructions were given to make the letter as persuasive, fresh, and original as they could. Time given to discuss and write the letter was 45 min. A similar task has been used by Fiedler (1967).

Chart task. Each group was given sheets showing the scores of Army apprentices in examinations held at various apprentice schools. They were required to use this information to construct two charts showing the results for two different years. A sample chart was provided and also written instructions on how to calculate a percentage and construct a chart. Groups were asked to work as quickly and as accurately as they could. The time taken by different groups to complete the task varied, but average time was 40 min. The task involved the separate subtasks of (a) counting the number of apprentices who passed, (b) calculating percentages of passers, and (c) constructing the chart. Two people were required to work separately on counting the number who passed, one person to calculate percentages, and the fourth person to construct the chart itself.

The structures of the work organizations used for these two group tasks are shown in Figure 1. In these graphs each position and subtask are represented by points. The allocation relationships are represented by directed lines from positions to

tasks, and the precedence relationships by directed lines from subtask to subtask. The collaboration and coordination values are given in Table 1.

Ability. Upon entry to the Army, each soldier had been administered an Army General Classification Test (AGC). This test is a group test which includes a variety of item types including analogies, number series, verbal reasoning, patterns, and circle series. This test was produced by S. Hammond and G. Bradshaw as a general classificatory test for IQ range 70–130. The test correlates .83 with the Otis Intermediate, .76 with the Otis Higher, and .78 with Raven's progressive matrices test.

Productivity criteria. The letters produced were rated by six judges who were all psychologists or graduate students in psychology. None of the raters was responsible for the design of the study. Two of them were full-time army officers. Each rater was given a short training period to acquaint him with the five dimensions on which each letter was to be judged. These dimensions were (a) well-written, clear versus poorly written, sloppy, awkward, (b) understandably presented versus confused, incomprehensible, (c) interesting versus boring (d) persuasive versus unconvincing, and (e) original, creative versus trite, platitudinous.

Ratings for each letter were summed over all judges using the procedure advocated by Cronbach, Gleser, and Rajaratnam (1963). Interrater reliability was .92. For the chart task, quality measures based on number of errors were obtained using two judges. Interrater reliability was .95. Performance scores on each task were converted to 50–10 modified standard scores.

RESULTS

Correlations between AGC scores and productivity were obtained for both sets of groups. These correlations are presented in Table 2. The correlation between the summed AGC score of a group and productivity was significant at the $p < .05$ level for the chart task, but not significant for the letter task.

Similarly, the correlation of the dullest and brightest man's AGC score with productivity was significant in the chart task groups, but not in the letter task. The correlation of the leader's AGC score with productivity was small and insignificant for both task groups.

It is apparent that the contributions of group members' abilities toward productivity depends on the group's task. When the task requires a high degree of collaboration, it appears that abilities of members are not related strongly to group productivity. However, when the task requires a high degree of coordination, abilities of members are related

TABLE 1

VALUES OF COLLABORATION AND COORDINATION FOR TASKS USED IN THE ARMY AND LABORATORY STUDIES

| Task | Collaboration | Coordination |
|---|---------------|--------------|
| Army study | | |
| Letter | 1.00 | 0.00 |
| Chart | 0.00 | .40 |
| Laboratory study | | |
| Story: coordination structure | 0.00 | .75 |
| Story: collaboration structure | 1.00 | 0.00 |
| Story: coordination-collaboration structure | .33 | .55 |

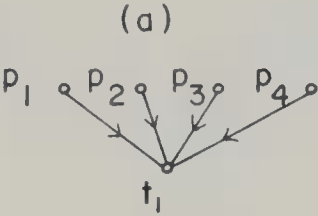
strongly to group productivity. Although the results obtained are consistent with predictions made concerning the effect of task structure on ability-productivity correlations, it is possible that the results could be interpreted in terms of different abilities required by the two tasks. Perhaps the abilities measured by the AGC score were relevant to the chart task only. Hence, an appropriate way to support the structural interpretation of the results would be to give a number of groups the same task or goal but vary the work organizations required to complete the task.

Study II—Laboratory Study

This study was designed to study the effects of organizational structure, leadership style, and member compatibility on small group creativity (Ilgen & O'Brien, 1968; O'Brien & Ilgen, 1968). Three kinds of interacting organizations were employed which differed in the amount of cooperation required. The goal was to construct three stories from three TAT pictures. Sixteen three-man groups were formed for each organization. These groups were matched in leadership style (as measured by Fiedler's LPC scale) of the appointed leader and the personal compatibility of group members (as measured by Schutz's FIRO-B scale) (Fiedler, 1967; Schutz, 1958). American College of Testing (ACT) scores on English were available for each S.

Work organizations. Organization 1: Coordination, but no collaboration. Each member

Army Study



Letter Task:
collaboration structure

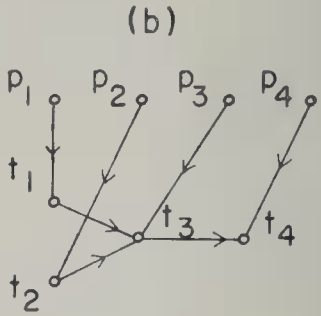
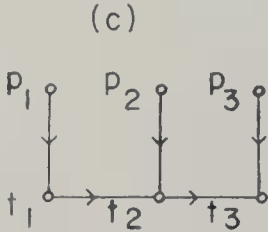
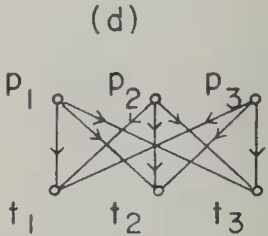


Chart Task:
coordination structure

Laboratory Study

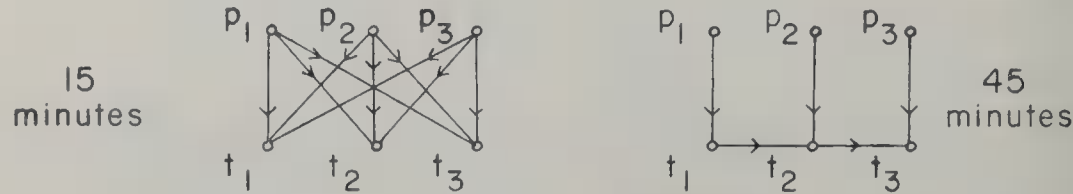


Story Task:
coordination structure



Story Task:
collaboration structure

(e)



Story Task:
collaboration-coordination structure

FIG. 1. Digraphs showing the organizational structure employed by groups in the army and laboratory studies. Directed lines show the allocation relationships between positions and tasks ($\overrightarrow{p_i t_j}$) and the precedence relationships between tasks ($\overrightarrow{t_i t_j}$).

started working on one story and after 20 min. passed his story on to the next man and received a story already started by the third man. After another 20 min., another exchange was made. In this manner all members worked on each story, but not at the same time. Organization 2: Collaboration, but no co-

ordination. All members worked together on each story for 60 min. Organization 3: Collaboration and coordination. Members worked together on all stories for 15 min. and then followed Organization 1. Digraphs showing the structure of these organizations are given in Figure 1.

TABLE 2
CORRELATION (PEARSON'S r) OF MEMBER ABILITY SCORE WITH GROUP PRODUCTIVITY

| Correlation of group productivity with | Tasks | | | | |
|--|--------------------|--------------------|--------------------|--------------------|----------------------------|
| | Army study | | Laboratory study | | |
| | Letter | Chart | Coordination | Collaboration | Coordination-collaboration |
| Sum of group abilities | .13 ($N=20$) | .58* ($N=20$) | .52* ($N=16$) | .03 ($N=16$) | .52* ($N=16$) |
| Ability of dullest group member | .12 ($N=20$) | .56* ($N=20$) | .49* ($N=16$) | -.04 ($N=16$) | .56* ($N=16$) |
| Ability of brightest group member | .12 ($N=20$) | .48* ($N=20$) | .32 ($N=16$) | .15 ($N=16$) | .19 ($N=16$) |
| Ability of group leader | -.04 ($N=20$) | -.05 ($N=20$) | .41 ($N=16$) | .15 ($N=16$) | .25 ($N=16$) |

Note.— N = number of groups used in calculating correlation.
* $p < .05$.

Productivity criteria. The stories were rated by five graduate students of English on plot originality, elaboration, plot structure, sentence structure, expressiveness, humor, and suspense. Interrater reliability was .82 using the Spearman-Brown correction.

RESULTS

Correlations between summed ACT English scores and productivity for summed, brightest, dullest, and leader scores are shown in Table 2. The dullest and summed measures were the only significant correlations, and these occurred only in the organizations which required coordination. Hence, the results obtained in this study are consistent with those in the Army study in that task-relevant abilities were significantly related to group productivity only in those task organizations requiring coordination and then only for the summed abilities and the abilities of the dullest member in each group.

DISCUSSION

The significance of these results lies in the demonstration that the contribution of member intelligence to group productivity is dependent on both the ability of the member and the kind of task organization employed. In tasks where there is a high degree of collaboration, it appears that members are unable to contribute significantly because the organization involves a great deal of interaction and prevents the group from organizing the best contributions in a systematic fashion. Some evidence to support this interpretation comes from observer ratings of group interaction. In the creativity study, observers recorded the number of comments made by each member and also the number of disagreements between members on the content of their stories. Collaborative organizations generated more comments and more disagreements than organizations requiring only co-

TABLE 3
NUMBER OF COMMENTS AND ARGUMENTS FOR THE THREE TASK ORGANIZATIONS
IN THE LABORATORY STUDY

| Number | Task organization | | |
|---|-------------------|---------------|----------------------------|
| | Coordination | Collaboration | Coordination-collaboration |
| Median number of comments made by group members during task performance | 72 | 547 | 224 |
| Mean number of arguments per 5 min. session | 1.25 | 3.94 | 2.56 |

ordination (Table 3). For the groups working with an entirely collaborative structure, high level of interaction was associated with significantly lower productivity (O'Brien & Ilgen, 1968). Organizations involving some degree of coordination had higher productivity and less interaction than collaborative organizations.

In a task where there is low collaboration but high coordination, each member must make some contribution to the formation of the group product. Under these conditions, it is not possible for a single person to make the only major contribution, but it is possible for the group to organize systematically the contributions of the group members. For a task of this kind, the principle, "A chain is only as strong as its weakest link," seems appropriate. Poor work by a relatively dull person may severely limit the performance of brighter members. Only when all members have high ability for their particular task is it possible for group performance to reach a maximal level.

The correlations between leader ability and group productivity are low and insignificant in both studies, although they tend to be higher in the laboratory study. These results suggest that leaders in such groups do not have a substantial direct effect on group performance. The magnitude of their contribution is probably dependent on both the structure of the task and the status structure within the group. The highest correlations occurred in laboratory where coordination was used and where there were only two rank levels (leader and member). The majority of army groups had more than two rank levels. Further research is needed to investigate systematically the interrelationship between leader abilities, status structure, and group task organization.

The results of these studies may be specific only to tasks where the group is required to combine various sources of information into one final product. Further research should be devoted to identifying organizational effects when the task requires groups to generate a large number of products from a limited num-

ber of resources (e.g., producing alternative solutions to a human relations problem). It may be that different task types require groups to have different organizations for optimal effectiveness. These results suggest also that the assignment of individuals to groups should be made after consideration of both their abilities, the ability of other group members, and the type of task organizations. It seems to be inefficient to assign members of high ability to groups where the task allocation relationships are such that their contributions are going to be limited by the poor performance of relatively incompetent members.

REFERENCES

- CRONBACH, L. J., GLESER, G., & RAJARATNAM, N. Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 1963, 16, 137-163.
- FIEDLER, F. E. *A theory of leader effectiveness*. New York: McGraw-Hill, 1967.
- GIBB, C. A. Leadership. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. 2 vols.
- HESLIN, R. Predicting group task effectiveness from member characteristics. *Psychological Bulletin*, 1964, 62, 248-256.
- ILGEN, D., & O'BRIEN, G. E. The effects of task organization and member compatibility, on leader-member relations in small groups. Technical Report No. 58, July, 1968, University of Illinois, Group Effectiveness Research Laboratory.
- MANN, R. D. A review of the relationship between personality and performance in small groups. *Psychological Bulletin*, 1959, 56, 244-270.
- O'BRIEN, G. E. The measurement of cooperation. *Organizational Behavior and Human Performance*, 1968, 3, 427-439.
- O'BRIEN, G. E., & ILGEN, D. Effects of organizational structure, leadership style, and member compatibility upon small group creativity. *Proceedings of the American Psychological Association*, 1968, 3, 555-556.
- OESER, O. A., & O'BRIEN, G. E. A mathematical model for structural role theory, III. The analysis of group tasks. *Human Relations*, 1967, 20, 83-97.
- SCHUTZ, W. C. *FIRO: A three-dimensional theory of interpersonal behavior*. New York: Rinehart, 1958.
- WITZ, K., & O'BRIEN, G. E. Indices of cooperation for use with work groups. *Journal of Mathematical Psychology*, in press.

(Received February 3, 1969)

ATTITUDES OF PROFESSIONAL APPRAISERS TOWARD AESTHETICS AND APPRAISAL ACTIVITIES¹

JACK M. WRIGHT²

AND

J. H. LEMLY

*American Institutes for Research
Palo Alto, California*

Georgia State College

An information questionnaire and Likert-type scales measuring attitudes toward the influence of aesthetic factors in realty appraisal were mailed to 100 industrial and 512 nonindustrial real estate appraisers throughout the United States. The purpose of the study was to determine if appraisers had a generalized attitude toward allowing aesthetic factors to influence their appraisal activities and, if so, to what extent they felt that it did in fact influence them. Results indicated that there were two relatively unrelated but approximately equally important attitudes toward the issue in question: One attitude reflected the concern of the appraiser for the intended users of the property (public vs. private concern or individuals) or the use for which the property was intended (recreational vs. business, etc.); the second attitude regarded concern for such programs as urban renewal, highway beautification, city planning, and modern architectural trends. These results also suggested that appraisers are relatively positive in their attitudes, and feel that more weight should be given aesthetics in appraisal than is being given or than they personally give. These attitudes were also found to be related to the age of the appraiser and the size of the city from which he operated.

Investments of time, talent, energy, and money by contemporary programs of urban renewal, highway beautification, city planning, etc., make extremely salient the relationship between the aesthetic aspects of land and land-based structures and the economic value of such property. Attitudes and opinions of the appraisers of such property constitute a reflection of the values of the larger society regarding this relationship. Such persons may also be considered to possess "informed opinions" regarding this matter, and, to some extent, constitute "opinion makers" for the larger society. Thus, their attitudes toward aesthetic factors as influences upon the appraised value of a property seem especially relevant to the general topic of the relationship between aesthetic and economic value. The focus of the present study was the assessment of the attitudes of realty appraisers

regarding the importance of aesthetic factors in appraisal activity. More specifically, the question being studied was the degree to which the aesthetic qualities of such structures and their surroundings should or could be allowed to influence their economic value, from the point of view of the real estate appraiser.

Aesthetics was conceptually defined in this study as the quality of sensory reaction to a physical or nonphysical phenomenon, giving the phenomenon some probability of evoking evaluative reactions in an observer. Such evaluative reactions may be considered emotional in nature and are based upon or reflect the conceptual system of the individual observer. These reactions are not strict determinants of overt behavior (such as appraisal or purchasing) but, in interaction with other factors, they guide and direct overt behavior. Such evaluative reactions are considered learned and therefore dependent upon the history and characteristics of the observer. Observer characteristics included in this study were age and sex. Principal factors in the background of the observer dealt with in this study were physical environment (e.g., geographic location), education, and socio-eco-

¹ This research was part of a larger project studying aesthetics and economics as they relate to land and land-based structures, and was supported by the State Highway Department of Georgia and the Bureau of Public Roads.

² Requests for reprints should be sent to Jack M. Wright, 1340 Claremont Drive, Boulder, Colorado 80302. This study was conducted while the author was at Georgia State College, Atlanta, Georgia.

nomic level, and involvement in such activities as urban renewal.

Factors other than the personal and social characteristics of the observer which are of relevance include such aspects of the property as its intended purpose and the type of use expected for the property. Thus, relevant features of the properties which were assessed were (a) whether it will be used for commercial, residential, or industrial purposes, (b) developed by public or private concern, or (c) used by individuals (e.g., family) or groups (e.g., corporations, city government).

Our expectations were that appraisers of industrial properties would have more neutral or negative attitudes toward taking aesthetic factors into consideration in appraisal activity and that appraisers of residential properties would be most positive in their attitudes. Other expectations included the following: Appraisers of higher socio-economic status and educational level and those residing in more cosmopolitan surroundings would be more positive in their attitudes than persons lower on these dimensions.

METHOD

Sampling. The population to be sampled included 2700 appraisers—all members of the American Institute of Real Estate Appraisers. These persons fell into two major groups: 100 industrial appraisers and 2600 nonindustrial appraisers. Questionnaires were mailed to all of the industrial appraisers and to 20% of the nonindustrial appraisers. The sample of nonindustrial appraisers was selected through stratified random sampling proportional to the number of appraisers living in each state in the United States and in the District of Columbia. The states of Hawaii and Alaska were left out of the sample for reasons of time.

Of the 512 questionnaires mailed to the nonindustrial appraisers, 326 were returned and 282 of these were usable. Of the questionnaires mailed to the industrial appraisers, 42 of the returns were usable. Thus, the final sample was composed of 12% of the nonindustrial appraisers and 42% of the industrial appraisers.

Sample characteristics. All except one of the respondents were male. Their age ranged from 33 to 81, with a mean age of 52 yr. and a standard deviation of 10 yr. Of 321 respondents, 140 had more than 16 yr. education, 181 had 14 to 16 yr., 56 had 12 to 14 yr., and only 2 had no college at all. Regarding income, 169 reported income from appraisal work of \$20,000 or less per year, 63 reported \$30,000 or less, and 84 reported above \$30,000. Of 322 persons reporting, appraisal was the full-time occupation of

139, it constituted 75% of the occupation of 51 persons, 50% for 58 respondents, and 25% or less for 74 respondents. The median for the size of the city from which these persons operated was 100,000 to 499,000 in population. Most of the sample, in their appraisal work, ranged across from one to three states; 31 worked only in one city; 147 worked in one state; 82 worked in two to three states; 43 worked in more than three states but not nationwide; 21 reported making appraisals throughout the nation.

Regarding involvement in activities which might be expected to influence their attitudes toward aesthetics, 11 persons were involved in condemnation proceedings, 2 were involved in urban renewal activities, 14 in highway or other government land acquisitions, and 108 in mortgage loan appraisal. Of the persons reporting, 147 were involved in two or more of these activities. Only 5 persons reported no involvement in any such activities.

Methods. Two instruments were mailed to Ss, with a letter of explanation. The first of these instruments was a general information questionnaire containing 12 questions designed to elicit information from them regarding personal and social characteristics. The second instrument contained 30 Likert-type statements designed to measure attitude toward the influence of aesthetic factors on reality appraisal. For the purpose of this questionnaire, aesthetics was defined for the respondent as being "*beauty or the appreciation of the beautiful*," after Cohen (1941). Also contained in this second instrument were 3 questions regarding actual behavior (as opposed to opinion or belief) of the respondent in his appraisal activity.

The response mode for the general information questionnaire varied from checking one of several alternatives to writing an answer. For the attitude statements, respondents encircled one of five alternatives to reflect the degree of agreement with the statement that they felt: Strongly agree, Agree, Undecided, Disagree, and Strongly disagree. These alternatives were weighted as 5 (Strongly agree) through 1 (Strongly disagree) for positive items; weights for negative items were reversed. The respondent's score was the sum of the weighted alternatives endorsed by him for these 30 items. High scores indicate a positive attitude toward allowing aesthetic factors to influence real estate appraisal.

Two of the last three items asked the respondent to encircle a number (with alternatives 0–10) to indicate how much weight he gave aesthetic factors in his last appraisal activity, and how much he generally gave such factors; 0 indicated very little or no weight and 10 indicated that aesthetic factors were the sole determinant of appraised value. The last question requested that the respondent indicate the nature of the last property appraised—residential, commercial, or industrial.

RESULTS

The 22 attitudinal items correlating with the sum score of the 30 items at a high level of statistical significance ($p < .01$) were se-

TABLE 1
OPINION INVENTORY

| Opinion | r |
|--|-----|
| 1A* Aesthetic qualities are more important in the appraisal of residential than of industrial property. | .37 |
| 2 Aesthetic qualities of a property have more weight in appraisal when the potential occupant or tenant is a private person than when the user is a group of persons, (corporation, city government, etc.). | .27 |
| 3B* I feel that the current prevalence of urban renewal programs indicates that greater emphasis must be placed on aesthetic factors in realty appraisal. | .44 |
| 4B* The federal highway beautification program is an indication of the greater demand by the public for beauty in land and land based structures. | .45 |
| 5B* Modern city planning is a testimony to the growing public desire to create an esthetically pleasing physical environment in which to live. | .40 |
| 6 Aesthetic factors are taken into consideration most of the time by most realty appraisers in their work. | .25 |
| 7]™ Much less consideration should be given aesthetic factors in realty appraisal than is presently given them. | .27 |
| 8 The public demand for efficiency, "modern appointments," and relative low cost is greater than its demand for aesthetics in single family housing. | .22 |
| 9* The suburban nightmare of "look alike" "cracker box" houses demonstrates a lack of concern for aesthetics on the part of the builders. | .34 |
| 10A* Aesthetics is more important in the appraisal of commercial than of industrial property. | .41 |
| 11A* Aesthetic factors are more important in the appraisal of residential than commercial property. | .33 |
| 12A* For property which is intended for recreational use, aesthetic factors must be given greater consideration in making an appraisal than for other properties. | .39 |
| 13 The beauty of a physical structure (e.g., a home) is more important than the condition of its site in affecting its appraised value. | .27 |
| 14B* The growing tendency in modern architecture to build homes which blend into their natural surroundings indicates an increased appreciation of the need to give more consideration to aesthetic factors in construction. | .37 |
| 15B* Most of my clients are actively concerned about the aesthetic aspects of the structures they wish to have appraised. | .34 |
| 16 I do not give aesthetic factors much weight in appraising property because I feel that the market gives much more interest to utilitarian or functional factors. | .26 |
| 17* Structures which possess historical features may reflect increased market value. | .31 |
| 18 Structures which possess traditional features may reflect increased market value. | .29 |
| 19 A well-maintained older residential property will usually increase in value. | .26 |
| 20 Aesthetic factors are more important in the appraisal of structures located well outside of the periphery of a city and its suburbs than in appraisal of structures which are more centrally located. | .24 |
| 21A* Corporate bodies such as city governments, etc., are less concerned with aesthetic features of properties they purchase than are private companies or individuals. | .31 |
| 22 Aesthetic factors are more important in appraisal of properties which are intended for commercial enterprises dealing in services than for those dealing in goods and production. | .30 |

Note.—Items marked with an asterisk (*) are those in the shorter 12-item version of this scale. Items marked with "A" belong to the first cluster. Items marked with "B" belong to the second cluster.

lected for data analysis; most items possessed correlations of .30 or more with total score. A split-half reliability coefficient (corrected) for these items was only .57, and analysis of interitem correlations indicated that two factors were contributing to total score. A second set of 12 items was selected on the basis of a correlation with total score of .30 or higher; the split-half reliability coefficient for this set of items was .75 (corrected). The 22 items originally selected are given in Table 1 along

with their correlations with total score; items included in the later 12-item scale are marked with an asterisk. The following report is based on the sum scores derived from the 22-item version of the scale.

For this 22-item scale, the theoretical range of possible scores is from 22 to 110 with an expected mean of 66. The actual mean obtained was 84.4 ($N = 325$), with a standard deviation of 7.1. Thus the distribution of scores was relatively skewed, with a greater

frequency of positive than of negative attitudes toward giving aesthetic factors weight in appraisal. On the whole, the attitudes sampled ranged from neutral to positive.

The attitude scores of respondents was then correlated with their report of the amount of weight given aesthetic factors in their last appraisal of a property. Attitude was found to be positively and highly significantly correlated with their report of the amount of weight given, a correlation of .33 being obtained with an N of 325 ($p < .001$). However, it is to be noted that the mean response on a continuum from 0 to 10 on this question was 3.0 with a standard deviation of only 2.2. The truncated distribution of scores on this question would be expected to reduce the size of the correlation obtained. Evidence of the validity of the scale rests on the content validity of the items, the correlations of items with total score, and the correlation of attitude score with reports of the respondents of actual weight given aesthetics by them in making their most recent appraisal.

Examination of the patterns of item inter-correlations and perusal of item content indicates that there are two relatively unrelated factors of approximately equal importance contributing to these scores. The first of these clusters of items (Cluster A) involves questions reflecting the degree of concern of the appraiser for the intended users of the property (public institutions such as city governments vs. private concerns such as companies or individuals) or for the use for which the property is intended (recreational vs. business, residential, or industrial use). The second cluster (Cluster B) involves questions regarding concern for such programs as urban renewal, highway beautification, city planning, modern architectural trends, etc. The pattern of correlations of these items with certain other key items ("Aesthetic factors are taken into consideration most of the time by most realty appraisers in their work" and "Much less consideration should be given aesthetic factors in realty appraisal than is presently given them") is highly consistent and seems to indicate the following: Many appraisers feel that greater weight should be given aesthetic factors in realty appraisal than is given them at present; further, they feel

that present concern with such programs as urban renewal, highway beautification, and city planning reflect a growing public desire for aesthetics to be given more weight. The specific items involved in these two clusters are identified in Table 1.

Regarding characteristics of the sample, age was found to be positively and significantly correlated with attitude score, such that older men were more positive in attitude toward giving aesthetic factors weight in assessing property ($p < .05$). Size of city from which the individual was operating was also significantly related to attitude scores. However, contrary to expectation, persons operating from smaller cities apparently were more positive in their attitudes than were persons operating from larger cities ($p < .01$). Other factors such as educational level, yearly income, and range of operation (how many states were covered in the individual's work) were found not to be related to attitude to a significant extent. This latter finding was surprising as it was expected that these three variables would reflect the general sophistication of the individual appraiser, which would in turn produce differences in attitude. Correlations involving age and city size improved when recomputed using the later 12-item scale score. However, correlations involving educational level, yearly income, and range of operation remained nonsignificant when recomputed using the scores from the 12-item scale.

The mean attitude scores of industrial and nonindustrial appraisers were also compared, but no differences were found; in fact, they were almost precisely the same, being 85.2 and 84.2, respectively. Both means had a standard deviation of 7.1. When compared on response to the question, "Most of my work is with: commercial, industrial, or residential property," the differences were in the expected direction but did not approach significance. By expected direction is meant that persons reporting that most of their work was with residential property had the highest mean attitude score, and those reporting that most of their work was with industrial property had the lowest mean attitude score; persons reporting that most of their activity involved commercial property had a mean attitude

score which was intermediate between the other two.

DISCUSSION

The question of whether real estate appraisers have a general attitude toward whether aesthetic aspects of a property should or do influence appraisal of that property was the focus of the present study. Results based on this sample indicate that they do. Further, their attitudes seem quite positive in this regard. However, the results also indicate that they give aesthetic factors less weight (by self-report) in actual appraisal activity than they appear to feel should be given to such factors.

There were found to be two general, unrelated, but roughly equivalent factors which were contributing to the attitude in question. One factor was composed of items regarding concern for such programs as urban renewal, highway beautification, city planning, etc.; evidence appeared to indicate that most appraisers felt that less weight was being given aesthetics in realty appraisal than should be given and that such programs were evidence of a growing public desire for aesthetics to be given more weight. The other factor included questions reflecting concern on the part of the appraiser for the intended user or the intended use of the property. Speculatively, these results taken together with the fact that appraisers appear to give aesthetic factors less weight than they feel should be given may indicate that the appraiser feels restricted or inhibited in actualizing his attitude by giving more weight to aesthetics. It may be that he feels that he is restricted thus by an inadequately developed public concern, or he may feel he is restricted by sources of power or influence that control both him and the public.

Although age of appraiser and size of city from which the appraiser operated were found to be related to the attitude in question, no stable inferences regarding the bases for these correlations can be generated. Size of city was negatively related with positivity of attitude, with persons from smaller cities being more positive. Among other possibilities, this finding may suggest that land and land-based structures in larger cities are so economically valuable as to preclude greater consideration of aesthetic values. Or it may reflect some aspect of the personality of the individual who seeks a smaller city from which to operate. The positive correlation of age with positivity of attitude may indicate a general "mellowing" and accompanying change in values, it may reflect the effects of longer experience in the profession, or it may indicate a generational difference.

It was expected that such factors as educational level, yearly income, and range of operation in the person's work might reflect some factor such as sophistication which would result in the holding of a different attitude. However, these factors were found to be unrelated to attitude. At least for the variable of educational level, the failure to generate a relationship may be the result of an overly crude measurement of the education variable; only four categories of response were provided for this variable.

Although in the expected direction, no significant differences were found between persons involved in residential, commercial, or industrial appraisal.

REFERENCES

- COHEN, J. A scale for the measurement of attitude toward the aesthetic value. *Journal of Psychology*, 1941, 12, 75-79.

(Received February 14, 1969)

DISTORTION OF DRIVERS' ESTIMATES OF AUTOMOBILE SPEED AS A FUNCTION OF SPEED ADAPTATION

FRANK SCHMIDT AND JOSEPH TIFFIN¹

Purdue University

Ten young male drivers were required to make four estimates of 40 mph speed after varying amounts of exposure to an adapting speed of 70 mph. The influence of these varying amounts of exposure to the adapting speed on speed judgments was studied, and a significant ($p < .01$) upward distortion of estimations of 40 mph was found to occur as a function of exposure to the adapting speed. The eta between treatment conditions and speed estimates was .72, and r was .71. These results were discussed in relation to their implications for accident rates and highway construction.

After driving at a constant speed for a period of time, drivers often report that this speed seems slower than it did at the beginning and that speeds slower than this level seem to be extremely slow. This phenomenon has been reported to occur immediately after drivers exit from a high-speed freeway or expressway onto secondary roads with lower speed limits, and has been suggested as a factor in accidents at such locations (Matsen, Smith, & Hurd, 1955, p. 24). It is also reported to occur at points in cross-country routes at which the driver is required to reduce his rate of speed while remaining on the same roadway, for example, while driving through small towns, roadside residential areas, school zones, etc. Speed adaptation may be an important factor in determining driver estimation of speed in many driving situations.

Many factors have been found to play a role in determining driver behavior. Goldstein (1961; 1967) has reviewed studies in which human characteristics have been related to driving behavior, and additional studies have been done since his most recent review (Barrett & Thornton, 1968; Schuman, Pelz, Ehrlich, & Selzer, 1967; Schwenk, 1967). Most research in this area is concerned with relationships between personality and perceptual variables and accident rates. These personality variables may also affect drivers' estimation of automobile speed, including the size of the car, its "noisiness," and the kind of experience which the observer used as a

basis of estimate. In this study no attempt was made to control for personality variables, but such factors as size and "noisiness" of the test car were held constant, so that each *S* served as his own control within his series of speed estimations.

As to the accuracy of *Ss'* estimates of speed there is some disagreement, with the older studies (Forbes, 1932; Richardson, 1916) finding *Ss'* estimates highly unreliable and inaccurate, and the more recent researchers finding them more accurate (Barch, 1958; Desrosiers, 1962; Olson, Wachler, & Bauer, 1961; Suhr, Lourer, & Allgaier, 1958; Weisman, 1964). Perhaps the older findings were affected by the relative unfamiliarity of *Ss* with automobiles, and by the nature of the then-existing automobiles and roads. Another factor contributing to these differences could be the fact that in the two older studies, *Ss* were not passengers or drivers in the automobile whose speed they were estimating, whereas of the more recent studies cited above, all except the Desrosiers (1962) study had *Ss* either driving the test car or riding in it as passengers.

Suhr, Lourer, and Allgaier (1958), Weisman (1964), and Barch (1958) found that drivers estimate speeds rather accurately in the 35-45 mph range. Errors in estimation in this speed range are described as slight, for example, 1-3 mph. Following these findings, the present *Es* selected 40 mph as the target speed to be estimated and attempted to demonstrate the effects of speed adaptation on estimates of this speed.

As noted by Denton (1966), speed adapta-

¹ Requests for reprints should be sent to Joseph Tiffin, Department of Psychology, Purdue University, West Lafayette, Indiana 47907.

tion has never been demonstrated outside the laboratory. Denton (1966) has developed a power function with a variable exponent to describe *S*'s estimates of speed in a simulated driving situation. This power function predicts the speed adaptation effect that it is the purpose of this study to demonstrate.

A search of the literature reveals only one previous study in this area. Barch (1958) attempted to demonstrate speed adaptation but reported "no evidence for speed adaptation in the speed judgments made by drivers while decelerating under the conditions of these studies [Barch, 1958]." He suggested that the explanation for the failure to demonstrate adaptation effects may have been that speed adaptation requires longer periods of constant speed, constant speed higher than the rates used, or both longer periods and higher speeds. Barch's *S*s drove 20 mi. at 50 mph with an average of 1.62 min. between estimates. In a second experiment, Barch (1958) increased driving time between estimates to 8 min. by reducing the number of estimates from 13 to 6. Again the adapting speed was 50 mph. However, the data still did not reveal significant adaptation effects. It was the conclusion of the present *E*s that the demonstration of speed adaptation requires both a higher adapting speed and a longer period of driving at the adapting speed.

METHOD

Subjects

The *S*s were 10 male undergraduate students from introductory psychology courses, ranging from 18-20 in age and from 2-4 yr. in driving experience. Average yearly mileage for the past 2 yr. ranged from 8,000 to 25,000 miles per year (mpy) with the median at 13,000 mpy. Each had obtained his license to drive at the age of 16. None was a professional driver, though one had driven a delivery truck for a time.

Apparatus

The automobile employed was a 1962 Chevrolet two-door hardtop equipped with a manual transmission and a six-cylinder engine. The car registered 59,000 mi. on the odometer. The regular speedometer was disconnected and the regular speedometer cable removed. A special speedometer cable, longer than standard, was installed and connected to a Stewart-Warner speedometer. This experimental speedometer was mounted inside a small cardboard box, allowing *E*, who rode in the front seat op-

posite *S*, to permit *S* to read his speed or not as the situation required, simply by positioning the box at the appropriate angle. A large red and white sign reading "Caution" was mounted on the rear of the test car for safety purposes.

The main test area was a 50-mi. concrete four-lane, divided, limited access highway with a speed limit of 70 mph. The secondary test area was a level half-mile section of secondary asphalt road.

Procedure

Before entering the test car each *S* was given a copy of the following instructions to read:

This is an experiment to see how well the average driver can estimate the speed of the automobile he is driving. You will be asked to drive over a predetermined course and to make four estimates of your speed. You will be told at what speed to drive between estimates and will be allowed to use the speedometer to aid in setting the requested speeds. Estimates will be made in the following manner: As you drive along at the requested speed, *E* will ask you to adjust your speed to a certain level (target speed) without aid of the speedometer. If you must decelerate to reach the target speed, do not use the brakes. Merely take your foot off the accelerator and allow the automobile to slow to the desired speed. When you feel that you have hit the desired speed, indicate this to *E* by saying the word "Now!" You should try to adjust to the target speed as fast as possible but accuracy is more important than speed. If you feel that the automobile has slowed to a rate of speed below the target speed, use the accelerator to coax the car up to what you feel is the requested speed.

Each *S* was then quizzed by *E* to assure that he understood the instructions. There was a total of four estimates per *S*. Each *S* was tested separately, with *S* driving during the actual speed estimations and *E* driving the distance between the two test sites (5.5 mi.).

Judgment 1. The first part of the study required each *S* to accelerate from a dead stop and to estimate 40 mph by saying "Now" when he judged the test car to be traveling at that speed. During this time the speedometer was visible to *E*, who recorded the estimates to the nearest mile, but not to *S*. After Judgment 1, *S* and *E* changed places and *E* drove to the main test area at 48-55 mph.

Judgment 2. At the main test area, the test car was stopped for 4 min. on the roadside, during which time *S* was instructed as to what to do at the end of the 4 min. At the end of the waiting period, *S* accelerated to 70 mph, held that speed for 5 sec., and then dropped, on signal from *E*, to the speed *S* judged to be 40 mph.

Judgments 3 and 4. After Judgment 2, *S* was instructed to accelerate to 70 mph and to hold that speed until he received further instructions from *E*. *E* assisted *S* in maintaining 70 mph by requests to

TABLE 1
MEAN ACTUAL SPEEDS GIVEN AS ESTIMATES OF 40
MPH AND SDs OF ACTUAL SPEED ESTIMATES
WITHIN TRIALS

| Trials | Mean estimate | SD |
|--------|---------------|------|
| 1 | 41.4 | 5.24 |
| 2 | 44.5 | 5.94 |
| 3 | 50.5 | 3.50 |
| 4 | 53.4 | 5.42 |

Note. $N = 10$, with each S giving one estimate on each trial.

increase or decrease speed. The tolerance range was 68–72 mph. (Additional requests were made near points of estimation in an attempt to keep the speed within the 69–71 range.) After 20 mi. (as measured by the odometer) at 70 mph, S was requested to drop to an estimated 40 mph. Then S accelerated to 70 mph again, maintained that speed for 20 more miles, and again made an estimate of 40 mph.

Design. Judgment 1 provided a measure of speed estimation under no adaptational effects or adaptation at 0 mph. Judgment 2 was a measure of the effects of minimal (5 sec.) exposure to 70 mph on the estimation of 40 mph. Judgment 3 was designed to furnish a measure of the effects of driving 70 mph for 20 mi. on the estimation of 40 mph, and Judgment 4 was a measure of the effects of driving at 70 mph for 40 miles. This last statement assumes, along with Barch (1958), that momentary decelerations from the adapting speed do not significantly reduce any ongoing adaptation process.

Analysis. A single factor repeated measures ANOVA was carried out on the data. Eta and r were also computed.

RESULTS

Table 1 presents the mean actual speeds given as estimates of 40 mph on each of the four trials, along with standard deviations within each trial.

Table 2 presents the results of the repeated measures ANOVA. The F ratio is highly significant. The Pearson r between treatment conditions and speed estimates was .71. Eta was .72.

DISCUSSION

As can be seen from Table 2, there was a very significant effect due to the speed adaptation treatment conditions. Although underestimation occurred under all treatment conditions, S s showed a strong tendency increasingly to underestimate their speeds as exposure to the 70 mph adapting speed increased. A

comparison of the magnitudes of eta (.72) and r (.71) shows that the relationship between speed estimates and treatment conditions is almost perfectly linear. Since no attempt was made to distribute the treatment conditions at equal intervals along scale, this linearity is entirely fortuitous. Eta is a better measure of the relationship of interest here.

In view of these highly significant results, Barch's (1958) speculation to the effect that his failure to demonstrate speed adaptation effects was due to his use of too short periods of constant speed and/or too low constant speeds is credible.

Questioned after the study, all S s reported previous knowledge of speed adaptation and all stated that they made conscious adjustments to offset adaptation effects. One S explained to E after the test, "You have to seem to be crawling to do 40 mph after driving at 70 mph." An important question is whether or not drivers make conscious adjustments for speed adaption in everyday driving. It may well be that S s in this study made adjustments because of a temporary desire, induced by the test situation, to be accurate in their estimates. This adjustment-inducing factor may not exist in day-to-day driving. The fact that there is a relatively high rate of accidents at the terminations of highways (Department of Scientific and Industrial Research, 1962) indicates that, if such adjustments are made in everyday driving, they are not complete or accurate enough to overcome the effects of speed adaptation. Speed adaptation may be impossible to measure precisely in its normal context. Something like the principle of indeterminacy in physics may be operative here: By the

TABLE 2
SUMMARY OF ANOVA

| Source of Variance | SS | df | MS | F |
|--------------------|--------|----|--------|--------|
| Between people | 196.3 | 9 | 21.81 | 13.05* |
| Within people | 1520.9 | 30 | 50.69 | |
| Treatments | 900.1 | 3 | 300.03 | |
| Residual | 620.8 | 27 | 22.99 | |
| Total | 1717.2 | 39 | | |

* $p < .01$ level.

very fact of measuring drivers' behavior, *E* may be inducing changes in that behavior. However, one need not measure this phenomenon that finely to demonstrate its existence, as has been shown in the present study.

Speed adaption as demonstrated in this study is a factor that highway and traffic engineers should take into account when designing certain parts of our transportation network, for example, exit and entrance ramps on high-speed roadways, curves at the end of long stretches of straight road, the setting of speed limits, etc. As more is learned about this phenomenon, specific recommendations will undoubtedly be formulated. At this point, however, the real need is for further research in this area.

REFERENCES

- BARCH, A. M. Judgments of speed on the open highway. *Journal of Applied Psychology*, 1958, 42, 362-366.
- BARRETT, G. V., & THORNTON, C. L. Relationship between perceptual style and driver reaction to an emergency situation. *Journal of Applied Psychology*, 1968, 52, 169-176.
- DENTON, G. G. A subjective scale of speed when driving a motor vehicle. *Ergonomics*, 1966, 9, 203-210.
- DEPARTMENT OF SCIENTIFIC AND INDUSTRIAL RESEARCH, 1962, Road Research 1961. *The Report of the Road Research Board with the Report of the Director of Road Research*. London: H.M.S.O. pp. 37-38, 1962.
- DESROSIERS, R. D. Speed estimation on residential streets. *Public Roads*, 1962, 32, 74-76.
- FORBES, T. W. Measuring drivers' reactions. *Personnel Journal*, 1932, 11, 111-119.
- GOLDSTEIN, L. G. Research in traffic accident prevention: An overview of research supported by the Public Health Service. United States Department of Health, Education, and Welfare. *Traffic Safety Research Review*, 1967, 11, 50-56.
- GOLDSTEIN, L. G. Research in human variables in safe motor vehicle operation: A correlational summary of predictor variables and criterion measures. The Driver Behavior Research Project, George Washington University, 1961.
- MATSON, T. M., SMITH, W. S., & HURD, F. W. *Traffic engineering*. New York: McGraw-Hill, 1955.
- OLSON, P. L., WACHSLER, R. A., & BAUER, H. J. Driver judgments of relative car velocities. *Journal of Applied Psychology*, 1961, 45, 161-164.
- RICHARDSON, F. E. Estimations of speeds of automobiles. *Psychological Bulletin*, 1916, 13, 72-73.
- SCHUMAN, S. H., PELZ, R. C., EHRLICH, N. J., & SELZER, M. L. Young male drivers: Impulse expression, accidents, and violations. *Journal of the American Medical Association*, 1967, 200, 1026-1030.
- SCHWENK, L. C. Personality correlates of accident involvement among young male drivers. *Dissertation Abstracts*, 1967, 27, 3734.
- SUHR, V. W., LOURER, A. R., & ALLGAIER, E. Judgments of speed on the highway and on the auto trainer. *Traffic Safety Research Review*, 1958, 12, 27-31.
- WEISMAN, A. D. Driver estimates of automobile speed. Unpublished manuscript, University of Southern California, Los Angeles, 1964.

(Received February 24, 1969)

EX-517.070

Marygrove College Library
Detroit, Michigan 48211
PLEASE DO NOT REMOVE

journal of applied psychology
monograph

Vol. 53, No. 1, Part 2

February 1969

**Personality, Problem-Solving Procedure,
and Performance in Small Groups**

By

Thomas J. Bouchard, Jr.

Editor: KENNETH E. CLARK

Published by the American Psychological Association

journal of applied psychology

monograph

Vol. 53, No. 1, Part 2

February 1969

PERSONALITY, PROBLEM-SOLVING PROCEDURE, AND PERFORMANCE IN SMALL GROUPS¹

THOMAS J. BOUCHARD, JR.²

University of California, Santa Barbara

Two experiments were conducted. The first compared a combination of group and individual brainstorming with simply individual brainstorming. The second contrasted three problem-solving procedures: critical group problem solving, group brainstorming, and individual brainstorming. In Experiment II, all three procedures were divided into feedback and nonfeedback conditions. Feedback consisted of having Ss listen to the first third of their performance and then continue to work on the remainder. Performance under all conditions was correlated with personality variables derived from the California Psychological Inventory (CPI), the Firo-B, the Myers-Briggs Type Indicator, a vocabulary test, and five factors derived from a factor analysis of the CPI. Experiment I indicated that there is no difference between a combination of group and individual brainstorming and simply individual brainstorming. Experiment II indicated that individual brainstorming is superior to group brainstorming which is superior to group critical problem solving. Feedback had no effect on performance within procedures. The CPI Sociability scale and the first factor of the CPI were shown to be consistently related to performance under group-problem-solving conditions.

During the last 15 years there has been a rapid growth of research interest in the facilitation of creative or original thinking, and most of this interest has focused on the individual (Barron, 1965; Golann, 1963; Mac-

Kinnon, 1962; Stein & Heinze, 1960; Taylor & Barron, 1963). Another area of potentially equal importance, but less rapid growth, has been that of creativity in groups. In spite of the tremendous surge of research on the small group (McGrath & Altman, 1966) few studies have focused on processes or procedures for the facilitation of creative or original thinking in problem-solving groups. In their review of studies contrasting the quality of group performance and individual performance, Lorge, Fox, Davitz, and Brenner (1958) excluded the consideration of group process as such. Kelley and Thibaut (1954), in their review of experiments on group problem solving and process, spent less than one page on the effects of formal group-problem-solving procedures, and cited no relevant studies. Recent reviews

¹ This investigation was supported in part by a United States Public Health Service Predoctoral Fellowship held by the author (1-F1-MH-23, 747-01A1) from the National Institute of Mental Health while at the Institute of Personality Assessment and Research, University of California, Berkeley. Part of the research presented here was carried out in partial fulfillment of the PhD degree at the University of California, Berkeley. The author expresses his appreciation to Donald W. MacKinnon, Gerald Mendelsohn, and William Rohwer for their advice and criticism.

² Requests for reprints should be sent to the author, Department of Psychology, University of California, Santa Barbara, California 93106.

by Hoffman (1965) and Maier (1967) cite a few relevant studies and deal with the question at some length.

Most of the studies that have been conducted have simply dealt with the contrast between the individual and the groups. This is clearly an inadequate formulation of the problem. Whether or not group problem solving is superior to individual problem solving, it is clear that group approaches to problems requiring creative solutions are becoming more and more necessary as the accumulation and fractionation of knowledge increases. Therefore, a question such as "Is individual brainstorming superior to group brainstorming?" is misplaced. Rather, one should ask such questions as "Under what conditions will which method solve which sort of problems?" "Which type of people work best using which methods?" "What is the optimum combination of these methods?" "What is the best combination of group and individual work?" In all of these cases, however, a "nominal" group made up of the nonoverlapping scores of an equal number of individuals who work on the same problem for the same amount of time can serve as a very useful base line for evaluating and understanding the effects of a particular technique. The studies to be presented in this monograph have attempted to answer some of the above questions. Four procedures, *group brainstorming*, a *combination of group and individual brainstorming*, *group critical problem solving*, and *individual brainstorming* have been studied in some detail and individual differences have been examined within procedures.

EXPERIMENTAL STUDIES OF GROUP-PROBLEM-SOLVING PROCEDURES

Experimental studies of group-problem-solving procedures have been reviewed a number of times (Hoffman, 1965; Salvatore, Willis, & MacKinnon, 1966). Here only those studies directly relevant to the experiments which follow will be discussed. These studies fall into three broad categories: (a) studies comparing group versus individual problem solving, (b) studies comparing different forms of group problem solving, (c) studies of brainstorming by individuals.

The first study comparing individual and group brainstorming was conducted by Taylor, Berry, and Block (1958). The study was designed to investigate "whether group participation when using brainstorming facilitates or inhibits creative thinking." They found that although the mean score (number of ideas) for 12 groups of four Ss each was higher than the mean score for 48 individuals, the mean score for 12 nominal groups (combined output of four randomly selected Ss who worked as individuals) was higher than the mean score of the 12 real groups. The nominal groups also produced more unique and higher quality responses than the real groups. Analysis of covariance showed that this was due almost entirely to the larger number of ideas. Taylor et al. (1958) concluded that "To the extent that the results can be generalized, it must be concluded that group participation when using brainstorming inhibits creative thinking [p. 43]."

Dunnette, Campbell, and Jaastad (1963) repeated Taylor et al.'s study using a modified design which allowed the same individual to participate in both the individual and group brainstorming sessions. The experiment was performed twice, using research scientists in one sample and advertising men in the other. The results showed that individual brainstorming is superior to group brainstorming in both quality and quantity of ideas and that the largest number of ideas is produced under individual brainstorming conditions after a group session. One of the major difficulties with this study is that a female *E* was present during the brainstorming session. This may have had an inhibiting effect over and above the fact that it was a group situation.

Parnes and Meadow (1963) have reported briefly on a study similar to the one reported by Taylor et al. (1958). They compared nominal groups of individuals who worked under conventional (critical thinking calling for quality solutions as opposed to quantity) procedures with real groups using deferred judgment. The real groups using deferred judgment were significantly more productive of good ideas than the nominal groups using critical problem solving. In a second com-

parison the individuals who made up the nominal groups used deferred judgment. In this case, the results failed to replicate Taylor et al. since there was no significant difference between the two conditions.

A study by Taylor and Block (1957) was designed to answer the question "Should group or individual work come first on problems requiring creative thinking when equal time is devoted to each?" Arguing that if the group work came first, a common set might reduce the variety and number of ideas produced in response to the problem, they divided 72 Ss into two groups and had them work on nine problems over a 4-day period. One group spent the first half of their time working on a problem alone and the other half working on the same problem in a group of three. The other half of the Ss followed the reverse procedure. The instructions emphasized critical thinking (quality) and quantity; they were not brainstorming instructions. It should be noted that all responses under both conditions were written and not tape-recorded as in the previous experiment (Taylor et al., 1958). The results indicated that it did not make any difference in number of ideas produced whether the individual or group work came first. The 12 groups of three, whose members worked alone first, were combined into nominal groups and compared with the 12 groups of three who had worked as real groups for the same period. The results showed that "Under these conditions, group participation inhibits rather than facilitates the production of ideas [Taylor & Block, p. iv]." These results corroborated the earlier study. They do differ, however, from those of Dunnette et al. (1963), who showed that, under brainstorming conditions, individual brainstorming is superior after, rather than before, group brainstorming.

Although Tuckman and Lorge (1962) do not deal with creative problem solving in the sense it is being used in this paper, they have introduced a paradigm that is applicable to the individual versus groups comparison under discussion. They argue that if a group is a more effective problem-solving unit than the individual, groups should generate a product better than the best ideas of its individual

members. To test this hypothesis, they had 70 individuals work on the Mined Road Problem alone, and then as five-man groups in a re-solving situation. As a control condition, 70 groups of five men also worked initially as groups. When a comparison of means was made, the re-solving groups performed better than their members had as individuals. However, when composite scores of the best ideas given under individual conditions in each of the groups were compared with their re-solving scores, the composite solutions were superior to the re-solving scores. A comparison between the re-solving groups and control groups that simply worked initially as a group yielded no significant difference suggesting no practice effect. Tuckman and Lorge concluded that "The group does not even incorporate or summate the best ideas of its members [p. 49]."

Another worthwhile variation on the problem of individual versus group comparison has been introduced by Campbell (1968). He had second- and third-line managers work on N. R. F. Maier's (1967) "Change of Work Procedure" problem under three conditions: (a) individual solutions, (b) individual solutions after a quasi brainstorming session, and (c) group consensus (four-man groups) in that order. No significant change in the average individual solution score was found following the quasi brainstorming session. Individual solutions combined into "nominal" group scores, using either average scores or a composite of the best elements for each of the four solutions, were significantly better than the real groups or average individual solutions.

We now turn to studies comparing various forms of group problem solving. Weisskopf-Joelson and Eliseo (1961) compared brainstorming instructions and critical instructions on such simple tasks as creating brand names for a cigar, a deodorant, and an automobile. Using seven persons per group (mixed sex), they compared three brainstorming and three critical groups. The brainstorming groups produced more responses (quantity), but the critical groups produced responses with a higher mean quality. Cumulative frequency distributions of the quality scores demon-

strate that at the high quality end of the distribution the two procedures do not differ in the number of responses produced. At the low quality end the brainstorming procedure produces more responses than the critical procedure.

Parloff and Handlon (1964) compared congenial and uncongenial dyads under brainstorming and critical conditions. Working on the hypothesis that perhaps brainstorming groups did not generate more good ideas than critical groups, but simply reported more of them, they had Ss write down their ideas and they also tape-recorded each session. They found that more good solutions were written under brainstorming conditions than critical conditions and that brainstorming groups generated (spoke and wrote) more solutions but not more good solutions than critical groups. Congeniality had no effect. They suggest "that the suspension of critical judgment may simply lower the subjects' standards for reporting ideas without substantially increasing their repertoire of "good" ideas [p. 25]."

This experiment differed from most of the others in one important respect. The Ss were made to work on the problem individually prior to their dealing with it in the group, in order to exhaust their store of answers. In this way, it was expected that any solutions generated would then be new and unique to the group. This may be true, but if one assumes that the active sharing of each other's ideas, whether they were part of the individual's repertoire prior to the interaction or not, is an essential aspect of brainstorming, then in a real sense this experiment fails to grasp an important aspect of the process itself. Along these same lines, it should be noted that the experimental units were female dyads. There is good reason to believe that dyads are a unique kind of group and generalizations about them to larger groups are tenuous (Thomas & Fink, 1963).

Brilhart and Jochem (1964) compared three different group-problem-solving procedures. The three different procedures used were as follows: Procedure I consisted of three steps: (a) analysis of the problem, (b) brainstorming using Osborn's instruction, and (c) setting up of standards and criteria to evaluate ideas. Procedure II was the same as I

except Steps *b* and *c* were reversed. Procedure III consisted of the following two steps: (a) analysis of the problem and (b) generation of solutions and evaluation of their relative merits at the same time (e.g., critical problem solving).

In terms of total number of ideas, Patterns I and II did not differ from each other, but they were both significantly superior to Pattern III. Using a more stringent criterion and eliminating ideas not rated above a level at which ideas were considered "good," Pattern I was better than Pattern III. But this difference was only marginally significant. The mode of recording responses in this experiment differed from most of the others. During each procedure a "recorder" wrote the gist of each idea on the blackboard. This mode is more similar to writing than it is to tape-recording.

Brilhart and Jochem concluded that patterns of problem solving which separate ideation and evaluation are superior to patterns which combine them. They comment that "evidently the emphasis usually placed on value and quality during a problem-solving discussion can dampen the expression of ideas of potential merit [p. 179]." This interpretation was also supported by Ss' responses to the questions "If you were to lead the discussion of a similar problem, which of the three sequences would you _____ most prefer to use, _____ least prefer to use?" Significantly more Ss chose I and III over II.

The following studies by Parnes and Meadow have evaluated the effects of brainstorming on individual creativity. All studies used a written mode for recording responses. Parnes and Meadow (1959) compared individuals under brainstorming and critical instructions. They found more good quality ideas under brainstorming instructions than under critical instructions. A second comparison between Ss trained in a creative problem-solving course emphasizing brainstorming and nontrained Ss, where both groups worked under brainstorming conditions, showed that the trained Ss produced a significantly greater number of good quality ideas than the untrained Ss. A third finding was a positive correlation between total quantity and number of "good" ideas produced

under all conditions. This was interpreted as suggesting that "The efficacy of brainstorming in producing an increment in good ideas is possibly the result of the increased quantity of ideas encouraged by the method [p. 176]." The question of the relationship between quantity and quality of ideas has also been discussed by Hyman (1960, reported by Parloff & Handlon, 1964) and Parloff and Handlon (1964). Both Parloff and Handlon and Hyman found that the proportion of "good" ideas did not vary with the number of ideas. In fact, the correlation between quantity and quality reported by Parnes and Meadow (1959) is to a large extent a statistical artifact due to the lack of independence between the two variables.

A second study by Meadow and Parnes (1959) was designed to evaluate training in the creative problem-solving course mentioned above. Under neutral testing conditions (Ss were not told to brainstorm) trained Ss were shown to be superior to a matched (IQ) control group, on a variety of measures. A further study indicated that the effects of this course persisted for at least 8 mo. (Parnes & Meadow, 1960).

Meadow, Parnes, and Reese (1959) compared the effects of brainstorming and critical instructions on individual problem solving. Four groups of eight Ss worked (individually) on one problem under each condition in a counterbalanced design. They found significantly more "good" solutions (rated in terms of uniqueness and value) under the brainstorming instructions. They also found more good solutions under brainstorming instructions when they came before critical instructions, than after.

PERSONALITY AND GROUP PERFORMANCE

This section deals with the interaction between personality and performance in small groups. This relationship has been dealt with in two complementary ways. The first approach assumes that certain personality characteristics or personality syndromes facilitate effective group performance, and the larger the number of group members who share these characteristics, the less likelihood of interpersonal conflict and the more effective the group is likely to be, given a fixed

level of ability (Cattell, Saunders, & Stice, 1953; Fiedler, Meuwese, & Oonk, 1961; Grace, 1954; Haythorn, 1953; Schutz, 1958).

The second approach favors heterogeneous rather than homogeneous groups, but the nature of the heterogeneous groups varies widely. Hoffman and Maier (1961) argue that heterogeneous groups defined by low personality profile correlations (nonspecific heterogeneity) tend to produce members with substantially different perspectives on the problem, thereby increasing the probability of good solutions or more solutions. Ghiselli and Lodahl (1958), on the other hand, found that a skewness measure which reflected the fact that one S in the group was considerably more dominant than the next most dominant member correlated higher with group performance than the average amount of a trait possessed by the group or the amount possessed by the highest scorer (Hoffman, 1959; Hoffman & Clagett, 1960; Pelz, 1956).

The mixed results found in these studies indicate that there is perhaps some value in both of the procedures. However, these results also indicate that perhaps the problem has been inadequately conceptualized. Clearly if the one approach assumes that certain personality syndromes facilitate effective group performance, and that all members should share this syndrome, then gross profile correlations or simple summation scores across unselected traits are inadequate to reflect or operationalize the hypothesis. The first step in an adequate conceptualization of the problem requires knowledge of the personality characteristics of effective group problem solvers. The second step requires experimentation with both types of groups, homogeneous and heterogeneous, with respect to these particular characteristics. One should not forget the very strong possibility that personality type may interact with both problem type and group-problem-solving procedure. For example, a shy, submissive, cautious, and methodical person is unlikely to be as effective in a brainstorming group dealing with a human relations problem as he would be in a highly structured group dealing with strictly objective types of problems.

With regard to the first step mentioned above, Mann (1959) and Heslin (1964) have

reviewed a number of studies relating a variety of personality measures to various small-group performance measures in an attempt to delineate the characteristics of effective group workers, but success along these lines, although encouraging, has been limited. The conclusion of both reviewers is that the best predictors of an individual's performance in a group are ability (general and specific) and general adjustment. Mann (1959) reports that "In no case is the median correlation between an aspect of personality covered here and performance higher than .25 and most of the median correlations are closer to .15 [p. 266]." Heslin (1964) who reports somewhat higher individual relationships concludes that "The relations of six different personality categories to group performance has been reviewed. The direction of the relationships has usually been clearly indicated, but these relationships are weak for predictive purposes [p. 254]." The generality of these conclusions and the low correlations reported limit the usefulness of this information with respect to the assignment of talent and points out that more precise predictive procedures are necessary if effective problem-solving groups are to be pre-selected.

The task of bringing the relationships reported above to a level of more practical and perhaps theoretical importance can be approached in a number of ways. First, characteristics of other groups' members and the group structure can be manipulated in such a way that only certain types of personalities can perform effectively. Examples of this approach are studies which vary the social characteristics of other members (Breer, 1960), role requirements (Smelser, 1958; Speisman and Moos, 1962), and communication networks (Leavitt, 1951). This approach increases our ability to predict performance in one situation by decreasing the range of behavior which we attempt to predict. Thus an increase in precision is accompanied by a loss in generality and potential usefulness of the information at hand. A second approach consists of increasing the complexity of the predictor variables (Cattell & Stice, 1954). The use of a moderately complex predictor to predict a complex cri-

terion (productivity in a group) could perhaps increase both the generality and usefulness of the obtained measures.

In order to conduct any of the four kinds of studies mentioned above (studies of homogeneity, heterogeneity, manipulation of group structure, or complex predictors), more systematic knowledge of the relationship between personality variables and performance under a variety of systematically manipulated conditions would be very useful.

More generally, however, the motivation for studying personality variables in the following experiments reflects Cronbach's (1957) view that "Ultimately we should design treatments, not to fit the average person, but to fit groups of students with particular aptitude patterns. Conversely we should seek out the aptitudes which correspond to (interact with) modifiable aspects of the treatment [p. 681]."

EXPERIMENT I

Taylor et al. (1958) and Dunnette et al. (1963) established that real brainstorming groups are inferior to "nominal" brainstorming groups. The Dunnette et al. data suggest that an optimal order for combining individual and group work would be group problem solving followed by individual work. This suggestion was based on data obtained from different problems under each of the conditions. This excluded any test of the possibility that sets established during group discussion might carry over into individual sessions and limit the range of ideas considered. Elsewhere, Dunnette (1964) has suggested that the group section of a combination of group and individual problem solving be restricted "almost exclusively to a sharing of information" and that "ideas or suggested solutions to the problem should be scrupulously avoided." Whether this procedure would avoid the problem of sets is problematic, but it does insure that all relevant information will be considered. For many problems this procedure reduces simply to individual problem solving, the group session being useful, as Dunnette points out, only as a means of keeping people informed about decisions affecting them.

The Campbell (1968) and Tuckman and Lorge (1962) studies suggest that the order, individual problem solving followed by group problem solving, has no particular value. Also, the Taylor and Block (1957) study which used nonbrainstorming conditions found no order effect when both the groups and individuals worked on the same problems, and found very little evidence for set effects. If one assumes that the participants in a brainstorming group are exposed to information that they would not have had access to had they not been in the group, and that they can make use of this information when they work individually, then in conjunction with the above findings there is reason to believe that a combination of group and individual work, in that order, under brainstorming conditions, may be superior to simply individual work over the same period of time. The main purpose of this experiment is to test the above hypothesis. The author's brainstorming procedure differs in one major respect from previous studies. Under individual conditions, Ss wrote down their responses to the problem, rather than verbalized them.

Method

Subjects

The Ss were 48 male junior and senior students from an upper division psychology course. Participation in 2 hr. of experimentation was a course requirement.

Design

The experimental design which is similar to that used by Dunnette et al. (1963) is shown in Table 1. As in that study conditions, order and problem sets were counterbalanced. II indicates two consecutive 10-min. sessions of individual work. GI indicates 10 min. of group work followed by 10 min. of individual work.

Procedure

The experimental procedure was as follows: Ss met as a group and the rules of brainstorming were discussed. The instructions were:

This is an experimental study of brainstorming. You may not be familiar with this concept so I will give you a brief description of what it is and then discuss the rules with you. Essentially, brainstorming is a form of group interaction which is used to facilitate the flow of ideas. It is a technique widely used in a large number

TABLE 1
DESIGN OF EXPERIMENT I

| Group | Individuals | Order | |
|-------|----------------|-------|--------|
| | | First | Second |
| A | 1, 2, 3, 4 | II, 1 | GI, 2 |
| B | 5, 6, 7, 8 | | |
| C | 9, 10, 11, 12 | | |
| D | 13, 14, 15, 16 | GI, 1 | II, 2 |
| E | 17, 18, 19, 20 | | |
| F | 21, 22, 23, 24 | | |
| G | 25, 26, 27, 28 | II, 2 | GI, 1 |
| H | 29, 30, 31, 32 | | |
| I | 33, 34, 35, 36 | | |
| J | 37, 38, 39, 40 | GI, 2 | II, 1 |
| K | 41, 42, 43, 44 | | |
| L | 45, 46, 47, 48 | | |

Note.—1 = Problem Set 1, Thumbs and Education; 2 = Problem Set 2, People and Tourists; G = group; I = individual.

of U. S. corporations. It is generally used when new, unique, original, and creative ideas are desired. It is not used to solve everyday problems. The rules of brainstorming are straightforward and easy to comprehend. (1) Criticism is ruled out: Adverse judgement of ideas must be withheld until later. (2) Freewheeling is welcome: The wilder the idea the better. It is easier to tame down than to think up. (3) Quantity is wanted: The greater the number of ideas, the more likelihood of winners. (4) Combination and improvement are sought: In addition to contributing ideas of their own, participants should suggest how ideas of others can be turned into better ideas. Are there any questions?

The instructions were generally followed by a number of questions and the rules were elaborated. A demonstration tape of brainstorming was then played. It was an unrehearsed discussion by four male graduate students and dealt with the question "What would be the consequences if suddenly everyone could read everyone else's mind?" and it contained many highly original and unconventional ideas. The tape served to illustrate the range and type of ideas a brainstorming session could elicit. The Ss then dealt with the following two questions in order to get them used to the procedure and to each other. (a) Name as many uses as you can for a red brick. (b) Name as many uses as you can for a wire coat hanger. Each of these problems was dealt with for about 2 min. Following this the rules of brainstorming were repeated. Half of the groups then worked under a combination of group and individual work (GI) first. Their time was divided into 10 min. of group

brainstorming and 10 min. of individual brainstorming. Generally most of the groups' ideas had been expressed at the end of 10 min.; at this time they were given a copy of the problem and told to relax, read it over and see if they could think of any more new ideas or solutions on their own (this took place in a separate room for each S). They were urged to be unafraid to duplicate the ideas which might have been given in the group if they had any uncertainty as to whether or not they had been given previously. The Ss worked alone for 10 min. and then were reassembled for Problem 2. The above procedure was then repeated. The combination of group and individual work was followed by individual work (II). The Ss were assigned to an individual room, given a copy of the experimental problem and a stack of lined paper headed by the instruction "Be brief. Your answers need not be complete sentences." They were then asked to apply the principles of brainstorming as individuals. They worked on two problems for 20 min. each.

The Ss who were to work in the II condition first proceeded immediately from the practice problems to the II condition and then on to the GI condition. All group sessions were tape-recorded. The same E served for all groups, and participated only when it was necessary to reprimand criticism (which seldom occurred).

The problems used in this study are the same as those used by Dunnette et al. (1963). They are

Thumbs problem. We do not think this is likely to happen, but imagine for a moment what would happen if everyone after 1966 had an extra thumb on each hand. This extra thumb will be built just as the present one is, but located on the other side of the hand. It faces inward, so that it can press against the fingers, just as the regular thumb does now. Here is the question, what practical benefits or difficulties will arise when people start having this extra thumb?

Education problem. Because of the rapidly increasing birthrate beginning in the 1940's, it is now clear that by 1970 public school enrollment will be very much greater than it is today. In fact, it has been estimated that if the student-teacher ratio were to be maintained at what it is today, 50% of all individuals graduating from college would have to be induced to enter teaching. Here is the question. What different steps might be taken to insure that schools will continue to provide instruction at least equal in effectiveness to that now provided?

People problem. Suppose that discoveries in physiology and nutrition have so affected the diet of American children over a period of 20 years that the average height of Americans at age 20 has about doubled. Comparative studies of the growth of children during the last 5 years indicate that the phenomenal change in stature is stabilized so that further increase is not ex-

pected. What would be the consequences? What adjustments would this situation require?

Tourists problem. Each year a great many American tourists go to visit Europe. But now suppose that our country wished to get many more European tourists to come to visit America during their vacations. What steps can you suggest that would get more European tourists to come to this country?

The assignment of Ss to conditions was done in a systematic fashion so that one group in each of the four orders was completed before any condition was repeated. The three groups in each order were also matched as closely as possible with regard to the time of day during which the experiment was conducted.

Results

All ideas produced during both the group and individual sessions for each problem were transcribed. From the transcriptions a master list of all the different ideas for each problem was constructed. This list was used to rate the quality of ideas. All protocols (group and individual) were inspected to delete duplications within any one session. The protocols obtained during the individual part of the GI sessions were inspected to delete any ideas which had been previously expressed in the group sessions. Then the four protocols for each group's II sessions were inspected and all the different ideas extracted. The first analysis consists of comparing the number of different ideas or solutions contributed under the II conditions with the number contributed under the GI conditions. The means for various problems and conditions are shown in Table 2.

A three-way analysis of variance (Winer, 1962, p. 554, Plan 9) indicated that the only significant effect was due to problem sets. The Thumbs and Education problems yielded

TABLE 2
MEAN TOTAL NUMBER OF DIFFERENT IDEAS AND/OR SOLUTIONS TO PROBLEMS BY Ss UNDER CONDITIONS OF INDIVIDUAL-INDIVIDUAL AND GROUP-INDIVIDUAL BRAINSTORMING

| Problems | Individual-individual | Group-individual |
|------------------------|-----------------------|------------------|
| Thumbs and people | 73.34 | 74.67 |
| Education and tourists | 61.33 | 56.58 |
| Total | 134.67 | 131.25 |

TABLE 3

MEAN TOTAL NUMBER OF "GOOD" IDEAS AND/OR SOLUTIONS PRODUCED UNDER CONDITIONS OF INDIVIDUAL-INDIVIDUAL AND GROUP-INDIVIDUAL BRAINSTORMING

| Problems | Individual-individual | Group-individual |
|------------------------|-----------------------|------------------|
| Thumbs and people | 37.00 | 30.75 |
| Education and tourists | 11.17 | 11.33 |
| Total | 48.17 | 42.08 |

fewer responses than the People and Tourist problems.

A second analysis was conducted to find out if the number of "good" answers produced under II or GI conditions was significantly different. The scales used for rating quality of ideas were the same as those used by Dunnette et al. (1963) and Taylor et al. (1958). The effectiveness scale was used to rate the responses to the real problems (Education and Tourists) and the probability scale was used to rate responses to the imaginary problems (Thumbs and People). These scales are presented in Appendix A. The author and a second judge rated all problems. The sum of the two ratings was used as an estimate of a response's quality. Interrater reliabilities were computed on approximately one-third of the responses to each problem.

TABLE 4

MEAN QUALITY RATINGS FOR IDEAS AND/OR SOLUTION PRODUCED UNDER CONDITIONS OF INDIVIDUAL-INDIVIDUAL AND GROUP-INDIVIDUAL BRAINSTORMING

| Problems | Individual-individual | Group-individual |
|------------------------|-----------------------|------------------|
| Thumbs and people | 4.31 | 3.86 |
| Education and tourists | 2.79 | 2.85 |
| Total | 7.10 | 6.71 |

They were .67 and .59 for the Education and Tourists problems, and .62 and .55 for the People and Thumbs problems. Although not high, these reliabilites are adequate and in the same range as those reported by Dunnette et al. (1963). Any response with a total rating of 5 or better (sum of two raters) was considered "good." The means for various problems and conditions are shown in Table 3. An analysis of variance indicated that the only significant effects were due to problem sets. The Thumbs and Education problems yielded more good responses than the People and Tourists problems.

The next analysis deals with the mean quality of ideas and/or solutions. The appropriate comparisons are shown in Tables 4 and 5.

TABLE 5

ANALYSIS OF VARIANCE: MEAN QUALITY RATINGS FOR IDEAS AND/OR SOLUTIONS TO PROBLEMS UNDER CONDITIONS OF INDIVIDUAL-INDIVIDUAL AND GROUP-INDIVIDUAL BRAINSTORMING

| Source | df | Total (both problems of each set) | | Thumbs and people problems | | Education and tourists problems | |
|---------------------|----|-----------------------------------|-------|----------------------------|-------|---------------------------------|--------|
| | | MS | F | MS | F | MS | F |
| Between individuals | | | | | | | |
| Order (O) | 1 | 2.43 | 4.50 | 1.21 | 8.07* | .22 | |
| S × C | 1 | .18 | | .14 | | .00 | |
| S × O × C | 1 | .31 | | .13 | | .82 | 1.86 |
| Error _b | 8 | .54 | | .15 | | .44 | |
| Within individuals | | | | | | | |
| Condition (C) | 1 | .91 | 8.27* | 1.21 | 6.05* | .02 | |
| C × O | 1 | .00 | | .12 | | .17 | 2.43 |
| Set (S) | 1 | 1.06 | 9.64* | .02 | | .78 | 11.14* |
| S × O | 1 | .60 | 5.45* | .00 | | .56 | 8.00* |
| Error _w | 8 | .11 | | .20 | | .07 | |

* $p < .05$.

The II condition was clearly superior to the GI condition when both problems of each set were combined. The set effect was due to higher mean quality scores for the Thumbs and Education problems than the People and Tourist problems (7.11 versus 6.69). The significant Set \times Order interaction shows that this difference is due to inferior performance on the People and Tourists problems under the order II-GI.

A separate analysis of the unreal (Thumbs-People) and real (Education-Tourists) problems of each set is informative. The overall condition effect is shown to be entirely a function of the superiority of the unreal problems under the II condition. The superiority of the order GI-II over the order II-GI (4.31 versus 3.86) yielded a significant order effect for the Thumbs and People problem. The significant Set effect for the Education and Tourist problems was due to the higher quality of responses to the Education problem (3.00 versus 2.64). The Set \times Order interaction indicates that this difference was due to inferior performance on the Tourist problem under the order II-GI.

Discussion

The most striking aspect of this experiment is the failure to find large differences between various conditions in terms of number of answers as reported by Taylor et al. (1958) and Dunnette et al. (1963). Since Taylor's population of Ss was comparable to the population sampled in this experiment, a comparison of some of the data was attempted. Table 6 contains these comparisons. The data consist of the mean total number of responses to each of three problems. The data in parentheses are taken from Taylor et al. (1958), page 34, Table 2.

The other data are from this experiment. Each mean represents the mean performance of three groups for the first 10 min. of performance under the appropriate condition. All data were taken from the column headed "First" in Table 1. For example, the mean for the real groups under Tourist problem (35.3) represents the mean performance for the first 10 min. of Groups J, K, L, under GI,2. Table 6 clearly shows that the differences between the real groups are very small,

TABLE 6
MEAN TOTAL NUMBER OF RESPONSES TO EACH PROBLEM BY REAL GROUPS AND NOMINAL GROUPS

| Condition | Problems | | | | | |
|----------------|----------|------|--------|------|-----------|------|
| | Tourists | | Thumbs | | Education | |
| Real groups | (38.4) | 35.3 | (41.3) | 41.0 | (32.6) | 27.0 |
| Nominal groups | (68.3) | 39.3 | (72.6) | 42.7 | (63.5) | 30.0 |

Note.—Data in parentheses from Taylor et al., 1958, Table 2.

indicating that the brainstorming sessions were essentially replications, while the differences between the nominal groups are very large. Since Taylor et al.'s conclusion that nominal groups are superior to real groups across a variety of measures depends on the difference in number of responses under each condition, the discrepancy in Table 6 is of importance. There are a number of possible reasons for the discrepancy, such as a different introductory procedure, a different time limit (the present author's 10 min. versus Taylor's 12 min.), or the fact that Ss of this study had a copy of the problem and Taylor's Ss did not. The most crucial procedural difference between the experiments seems to be that the individual Ss of the present experiment wrote their responses rather than verbalized them.

This difference in results and procedures parallels the difference between Taylor et al. (1958), Dunnette et al. (1963), and Parnes and Meadow (1963). Parnes and Meadow (1963) report that they had their Ss record their own ideas in pencil as they spoke them. Apparently the individual Ss wrote their answers also since this is normally the way the test which they used as a problem is administered. On the other hand, Taylor et al. (1958) and Dunnette et al. (1963) had their Ss verbalize their ideas which were recorded on tape. That this difference in procedure is not trivial is attested to by the findings of Horowitz and Newman (1964). They analyzed spoken and written responses to two different questions and found that "Spoken expression produces more material (words, phrases, sentences), more ideas and subordinate ideas, more ancillary ideas, communicative signals, and orientation signals . . . [p.

646].” For number of ideas alone it took 10 min. to write what took only 2 min. to speak. This difference is likely to vary depending on the type of problem dealt with, but it is of such a large order of magnitude that the direction is unlikely to be reversed. Horowitz and Newman (1964) report that one of the main reasons for the difference between conditions is that “To a great extent our subjects could not tolerate silence—even a 10 or 15 second break seemed to create an uneasiness which tended to be filled in . . . [p. 647].” Another factor of importance was the greater feeling of commitment on the part of Ss who wrote their responses. Writing seems to have a permanence to it which verbalization, even when tape-recorded, does not. It seems that verbalization under individual conditions may exhaust the individual’s response repertoire more rapidly and completely than writing, with prolonged silences acting as a stimulus to continue to verbalize. These considerations lead us to believe that the findings of Taylor et al. (1958) are not contradicted; nevertheless, it should be noted that they may hold only under the special conditions where *S* verbalizes his responses aloud when working alone. Had the groups of the present experiment written their responses, Taylor et al.’s (1958) results would most likely have been replicated. Since there are much larger differences in sample and time between this experiment and that of Dunnette et al. (1963), no specific comparisons can be made.

Given the specified conditions of this experiment, verbalization in the group and written individual responses, the results indicate that in terms of total number and number of “good” ideas and/or solutions to a problem, it does not matter whether Ss work as individuals or in a combination of group work followed by individual work. This finding holds for what we consider to be two qualitatively different types of problems, unreal or imaginary (Thumbs and People) and real or concrete problems (Education and Tourists). For the criterion mean quality, however, individual performance is superior to a combination of group-individual work on *unreal* problems.

There are a number of possible reasons why, contrary to our predictions, the combination of GI work was not shown to be superior to II work. One aspect of the problem-solving process postulated by many investigators, but not considered in this experiment, is the stage of incubation. It is quite possible that an incubation period, or simply a short period of time during which to think about the problem without any pressure to produce answers, between the first and last 10 min. of problem solving, might benefit the GI condition more than the II condition. A second possibility is that the participants in the brainstorming groups were so busy trying to think of ideas they scarcely listened to what others were saying and did not make optimal use of the available information. It is also possible that training would improve the performance of the groups more than that of the individuals. Experiment II will focus, in part, on the second possibility.

EXPERIMENT II

The evidence is quite clear that individual brainstorming is superior to group brainstorming. It is not clear however that group brainstorming is superior to traditional group problem solving. The Weisskopf-Joelson and Eliseo (1961) study suggests it is not. Parloff and Handlon (1964) report that it is, only if we consider written but not spoken responses. Brillhart and Jochem (1964) report only marginally significant results in favor of brainstorming while using a written mode of response.

One of the difficulties inherent in evaluating these studies is their failure to make a clear distinction between the two procedures and to distinguish between the task and social-emotional aspect of group problem solving. In the experiment reported below, critical group problem solving and brainstorming are contrasted. The instructions have been written with both of these problems in mind. A detailed analysis is presented in Bouchard (1967). Briefly, group critical problem solving is conceptualized as a procedure which assumes that the best or most relevant solutions to a problem should be sought in light of a well-defined

goal and criteria of quality and relevance. The major problem-solving heuristic used is similar to Newell, Shaw, and Simon's (1960) means-ends analysis. In this situation, group members help keep the discussion on track by criticizing poor and irrelevant ideas. Brainstorming, on the other hand, is conceptualized as a procedure which involves a lessening of critical judgment and use of primary process or a sheer associative type of idea generating mechanism. This process is supposedly facilitated by other group members who provide unique stimuli which constantly prod the associative mechanism, thereby generating more and better suggestions.

The basic difference between the two types of procedures, from a task perspective, is a difference in the type of problem-solving heuristic used.

We have attempted to avoid as many potential interpersonal or social-emotional problems as possible by specifying the instructions explicitly enough so that they are precluded. To the extent that we are successful any differences between procedures are a function of the problem-solving heuristics used rather than interpersonal problems. The instructions were written in light of Tuckman's (1965) developmental theory of small groups. Their effectiveness has been evaluated by having Ss fill out a questionnaire after finishing work under each procedure.

Another important consideration is the interaction between type of problem-solving procedure and interpersonal behavior. Given the group situation and the motives characteristic of people in a group-problem-solving situation, can either of the procedures be used efficiently? The group-problem-solving situation can be described as a set of "reciprocally contingent interactions." Jones and Thibaut (1958) describe this situation in terms of the problem of interpersonal perception.

In reciprocal contingency situations, the need for information is immediate, and it must be quickly processed since neither actor has much time to think about the preceding act before having to act himself. As a consequence of this "urgency" consideration, we suggest that much of the perceiver's attentive energy will be directed to his own future responses and not the stable characteristics of the

other. Thus the main moment-to-moment problem is not "What is he like?" but "What am I going to do next?" [p. 158].

Analogously in the group-problem-solving situation, the demand characteristics are such that the main moment-to-moment questions are not "What are they saying?" but "What am I going to say next?" and "What will they think of what I say?" This type of set or attention distribution is likely to interfere more with the brainstorming heuristic, which emphasizes the stimulus value of others' responses, than with critical problem solving. We can circumvent this problem by providing task-relevant feedback, during a period in which the demands for productivity are reduced. This helps to achieve an optimal distribution of attention and information input. There is also evidence that techniques which allow Ss to become familiar with each other's thinking, regardless of whether or not they agree with it, facilitate interpersonal communication (Triandis, 1960) and creativity under some conditions (Triandis, Hall, & Ewen, 1965). The task-relevant feedback technique meets these requirements. In this experiment each problem-solving procedure is divided into a feedback and nonfeedback condition. Feedback consists of having Ss listen to a taping of their first 5 min. of performance on a problem, and then allowing them to continue to work on the same problem.

In order to evaluate the relative effectiveness of group brainstorming and critical group problem solving, "nominal" brainstorming groups were also formed. On the basis of past research, the "nominal" brainstorming groups are expected to be superior to the brainstorming and critical problem-solving groups. The feedback condition should be superior to the nonfeedback condition.

No prediction is made for the brainstorming and critical problem-solving groups, as such, but the feedback condition for each should be superior to the nonfeedback condition. The brainstorming group should benefit from feedback more than the critical problem-solving group. There should be a main effect due to feedback across all procedures.

Method

Subjects

The Ss were 144 male students drawn predominantly from lower division psychology courses in which participation in 3 hr. of experimentation was a course requirement.

Design

The experimental design, a $3 \times 2 \times 2$ repeated-measures design is given in Table 7 (Winer, 1962, p. 337). Due to time limitations only two problems were used (Thumbs and Education). Since each group worked on both problems (C factor), the order of solving the problems was counterbalanced within each $A \times B$ level. Pilot work had demonstrated that the time of day during which problem-solving sessions occurred was an important factor. Therefore it was systematically counterbalanced within problem order and condition.

Procedure

The instructions for each of the procedures are as follows:

Critical problem-solving instructions: This is an experimental study of group problem solving. Most of you have never worked on a problem in this way, so I will go over the procedure with you. This technique is a form of group interaction which is used to facilitate the flow of ideas. It is widely used in a large number of United States corporations. It is generally used when new, unique, original, and creative ideas are desired. It is not used to solve everyday problems.

The procedure is relatively straightforward and easy to comprehend. The rules are as follows:

1. The problem is analyzed. You should ask yourselves, both alone and as a group: "What

are its implications?" "What is happening?" "Why?"

2. You should try to determine, and keep in mind, what kind of criterion a solution should meet to be worthwhile. What defines a good solution?

3. In light of these two points, you should try to come up with as many good ideas as you can. By good ideas I mean if someone asks you to defend or explain what your suggestion means, you should be able to do so. It should also be meaningful and useful, not trivial. Criticism is acceptable but it should be directed only at ideas, not people or style of delivery. It should be pertinent and meaningful.

4. Everyone should contribute to the problem. Don't hesitate to speak up. As well as contributing ideas, individuals should also discuss what lines they are thinking along. In other words, we want you to maximize the exchange of information among yourselves. For example, you might mention what areas seem to you to be most productive of solutions, what kinds of assumptions are valid, etc. Let others know what you are thinking. Don't be afraid to say what you think.

I am going to tape-record the problem-solving session and take a few notes. Don't let this distract you.

If the group was designated a nonfeedback group, the instructions continued. "You will work on a problem for 20 minutes. After that you will work on a second problem for 20 minutes."

If the group was designated as a feedback group, the instructions were:

After you have worked on a problem for 5 minutes, I will stop you and we will take a break. During this time I will replay the tape. This will give you some idea of how you are doing and give you a chance to digest the ideas already put forth. It may perhaps suggest more new ideas. After the tape has been replayed, you will work on the problem for 10 more minutes. After that we will work on a second problem in the same way.

After this, both groups were asked "Are there any questions?" If there were questions the instructions were appropriately elaborated. Following this, the instructions were read through once again, and the problem was read to the group. If any S requested it, the problem was reread during the problem-solving procedure.

Group brainstorming instructions. The instructions for group brainstorming were identical to those for critical group problem solving except Steps 1, 2, and 3 were replaced by:

1. Criticism is ruled out; adverse judgment of ideas must be withheld. No one should criticize anyone else's ideas.

2. Freewheeling is welcome; the wilder the idea the better. It is easier to tame down than

TABLE 7
EXPERIMENTAL DESIGN

| | | c ₁ | c ₂ |
|----------------|----------------|-----------------|-----------------|
| a ₁ | b ₁ | G ₁₁ | G ₁₁ |
| | b ₂ | G ₁₂ | G ₁₂ |
| a ₂ | b ₁ | G ₂₁ | G ₂₁ |
| | b ₂ | G ₂₂ | G ₂₂ |
| a ₃ | b ₁ | G ₃₁ | G ₃₁ |
| | b ₂ | G ₃₂ | G ₃₂ |

Note.—a₁ = Brainstorming; a₂ = Critical problem solving; a₃ = Individual brainstorming (nominal groups); b₁ = No feedback; b₂ = Feedback; c₁ = Thumbs problem; c₂ = Education problem; G = six groups.

to think up. Don't be afraid to say anything that comes to mind, the farther out the idea the better. This will stimulate more and better ideas.

3. Quantity is wanted; the greater the number of ideas, the more likelihood of winners. Come up with as many as you can.

4. Combination and improvement are sought. In addition to contributing ideas of their own, participants should suggest how ideas of others can be turned into better ideas, or how two or more ideas can be joined into still better ideas.

Individual brainstorming instructions. The introduction to the instructions for individual brainstorming were the same as for group brainstorming. The following instructions were given:

The following rules are for groups. You will be working alone; nevertheless, I want you to apply these rules as best you can while working on these problems. What we are interested in is whether or not an individual can brainstorm and how he does it. The rules are as follows:

1. Criticism is ruled out; adverse judgment of ideas must be withheld. This is clear for a group. For an individual, it means don't criticize any idea that comes to mind. Say everything you think of.

2. Freewheeling is welcome; the wilder the idea the better. It is easier to tame down than to think up. Don't be afraid to say anything that comes to mind. The farther out the idea the better. It will stimulate more and better ideas.

3. Quantity is wanted; the greater the number of ideas, the more likelihood of winners. Come up with as many as you can.

4. Combination and improvement are sought. In a group, subjects are told they should suggest how ideas of others can be joined into still better ideas. For an individual this means that you should be willing to change suggestions you have made. Don't be afraid to combine and improve on them.

5. In a group, subjects are told "Everyone should contribute to the problem. Don't hesitate to speak up. As well as contributing ideas, individuals should also discuss what lines they are thinking along. For example, you might mention what areas seem to you to be most productive of solutions. Let others know what you are thinking. Don't be afraid to say what you think." In other words, they are asked to maximize the exchange of information among themselves. For an individual this means don't be afraid to mention things that come to mind other than straightforward suggestions. Talk about what you're thinking about. Don't be afraid to speak up.

The rest of the instructions were the same as for group brainstorming, and varied according to whether or not there was feedback.

Note that in this experiment, as opposed to Experiment I, responses under the individual brain-

storming conditions were tape-recorded. The author was the *E* and was present during all group-problem-solving sessions. He was not present during individual problem solving. The *E* did not participate and interrupted only when the procedural rules were not followed. This seldom occurred.

After working on both problems, *Ss* were asked to fill out a questionnaire which is presented in full in Appendix B. The questionnaire attempted to assess how well the instructions succeeded in equating the groups in terms of perceived criticalness, cohesion, satisfaction, and effectiveness.

Scoring

Both group and individual problem-solving sessions were transcribed in toto. All protocols were then inspected by two judges who removed duplicate ideas within group or individual sessions, marked each remaining idea or suggestion, and cross-checked each other. This analysis yielded the criterion—total number of ideas. In order to obtain a quality criterion, all ideas were rated for goodness using the following scales.

For the Thumbs problem a practicality-importance scale (see Appendix A) was used. Experiment I as well as previous studies had used a probability scale to rate this problem (Dunnette et al., 1963; Taylor et al., 1958). Reliabilities were fairly low, however, and it was felt that the scale did not grasp the dimensions of practicality or value, and so it was not used.

For the Education problem the same effectiveness scale as was used in Experiment I was employed. The effectiveness scale, although better than the probability scale, had also yielded poor reliabilities in past studies. It was felt that this was due to the fact that many ideas were difficult to rate outside of the context in which they were presented. In order to avoid this problem and minimize *E* bias, the ideas were judged in the following way. All ideas were numbered on the original protocol, then the protocols of the various conditions were systematically mixed. Each judge then rated one idea per page, across all conditions, by writing each idea number and its value on a small slip of paper until all ideas had been rated. If reading comments prior to the idea facilitated the rating, the judge did this. One judge was aware of the experimental condition from which some of the protocols were taken but the second was not.

The sum of both judges' ratings was used as the criterion. The intraclass correlation coefficient was used to assess the reliability of the two judges (Winer, 1962, p. 124). The reliabilities for four samples of 50 ideas were .77, .55, .70, .74 for the Education problem and .89, .67, .72, .73 for the Thumbs problem. The reliabilities are higher than those previously reported for these problems and are in an acceptable range. In order to improve the criterion further and reduce possible *E* bias to a minimum, both judges, together, went over all ideas with a discrepancy of 2 or more and reconciled all the discrepancies. The sum of both judges'

ratings was taken as the index of the idea's value. A "good idea" was one with a value of 4 or more for the Thumbs problem and a value of 6 or more for the Education problem. These lower limits were selected because a stricter criterion would have yielded zero scores for a number of groups. In both cases the lower limits are at the median of the distribution for summed ratings of a large sample (about 500) of the various conditions. The Ss who brainstormed individually were combined into nominal groups by combining the scores of any four Ss who had worked at the same time of day, and deleting overlapping ideas.

Results

Before presenting the results of the experiment on the dependent variables, the effectiveness of the instructions by means of the questionnaire will be evaluated. The instructions were explicitly designed to equate the groups as much as possible in terms of cohesiveness, motivation, lack of conflict, and criticalness with respect to self. There were no significant differences between feedback and nonfeedback groups within the procedures on any questions; these conditions, therefore, were collapsed. There were significant differences between procedures on only a few questions; nevertheless, the absolute values of the ratings for each of the procedures are of some interest and the results for all questions are reported in Table 8.

If one conceives of cohesiveness in terms of attraction to the group (Cartwright & Zander, 1960, Ch. 3), Questions 1 and 2 tap this dimension from two different standpoints. Responses to Question 1 indicate that Ss in the critical problem-solving group enjoyed working in their group significantly more than the brainstorming Ss. This finding runs counter to the intuitive impression of the *E* and nonsystematic reports from other experiments (Taylor et al., 1958). Note, however, that in both groups the mean rating is favorable. In response to Question 2, which could be taken as a measure of commitment to the group or motivation to work, there was no difference between conditions but the means were very low. This seems to indicate that S's commitment to the group qua group was not very high. Responses to Question 3 indicate that Ss in the critical condition had a greater commitment to the task, from a self-oriented standpoint, than Ss in either of the other two conditions. The differences are small and their absolute magnitude indicates that in all conditions Ss were neutral rather than favorable or unfavorable. Participants in the critical procedure also enjoyed working on the problems more than either of the other two procedures (Question 4). Again the differences are small. Participants in both brainstorming and critical problem-

TABLE 8
MEANS AND STANDARD DEVIATIONS FOR QUESTIONNAIRE RATINGS BY ALL THREE PROCEDURES

| Question no. | Brainstorming | | <i>t</i> -test ^a 1 vs. 2 | Critical problem solving | | <i>t</i> -test 2 vs. 3 | Individual brainstorming | | <i>t</i> -test 3 vs. 1 |
|----------------|---------------|-----------|--|--------------------------|-----------|---------------------------|--------------------------|-----------|---------------------------|
| | <i>M</i> | <i>SD</i> | | <i>M</i> | <i>SD</i> | | <i>M</i> | <i>SD</i> | |
| 1 | 6.27 | 1.88 | .01 | 7.69 | 1.39 | .05 | na ^c | na | .01 |
| 2 | 3.33 | 2.44 | | 3.96 | 2.31 | | na | na | |
| 3 | 4.48 | 2.64 | .05 | 5.67 | 2.29 | | 4.50 | 2.34 | |
| 4 | 5.15 | 2.17 | | 6.23 | 2.17 | | 5.66 | 2.19 | |
| 5 | 5.42 | 2.04 | .05 | 5.75 | 2.22 | | na | na | |
| 6 | 5.52 | 1.97 | | 5.72 | 1.70 | .01 | na | na | |
| 7 ^b | 6.21 | 1.25 | .05 | 5.96 | 1.46 | | 6.54 | 1.56 | |
| 8 ^b | 6.04 | 1.27 | | 5.96 | 1.34 | | 6.46 | 2.04 | |
| 9 | 6.13 | 2.68 | .01 | 7.02 | 2.30 | | 4.65 | 2.27 | |
| 10 | 7.33 | 2.29 | | 7.29 | 2.41 | | na | na | |
| 11 | 7.48 | 1.87 | .01 | 7.06 | 2.33 | | na | na | |
| | | | | | | | | | |

^a All tests two-tailed.
^b Feedback Ss only, *N* = 24, in all other comparisons *N* = 48.
^c Not asked.

solving groups expressed the same degree of satisfaction with self or group performance (Questions 5 and 6). Again the means fell at a neutral point.

All Ss who worked under feedback conditions in each of the procedures agreed on their judgment that feedback helped, and there were no significant differences between groups' judgments on this point (Question 7). This is of interest since it will be shown later that feedback was actually a hindrance to the critical groups. Apparently the feedback also made Ss somewhat more comfortable (Question 8). The ratings of perceived effectiveness indicate that individuals in the critical groups felt that their procedure was more effective than did those who engaged in individual brainstorming (Question 9). Surprisingly, this is just the opposite of the actual results. The absolute value of the means indicates that the critical groups perceived their procedure as effective while the individual brainstormers perceived theirs as ineffective. The Ss in both the critical and brainstorming groups reported being rather nervous, but they do not differ in this respect (Question 10). Neither of the groups reported feeling that other group members were critical of them (Question 11), and they do not differ significantly on this question.

The results for the three criteria, total number of ideas, number of good ideas, and mean number of ideas are presented in Tables 9, 10, and 11. The comparison between procedures consists of 20 min. of work under the nonfeedback condition and 15 min. of work and 5 min. of listening under the feedback condition. The comparisons are therefore in terms of equal man hours on the task, regardless of how the time was used. A comparison involving an equal amount of work time on the task will be presented later.

Total Number of Ideas and Number of Good Ideas

For a total number of ideas and number of good ideas, the main effect for procedure is significant at the .01 level (Table 11).

TABLE 9
MEANS FOR MAIN EFFECTS—20 MINUTES OF WORK

| Condition | Total no. | No. good | Mean quality |
|--------------------------|-----------|----------|--------------|
| Brainstorming | 29.8 | 14.1 | 4.31 |
| Critical problem solving | 21.0 | 9.9 | 4.47 |
| Individual brainstorming | 47.0 | 21.7 | 4.40 |
| Feedback | 31.2 | 14.6 | 4.45 |
| No feedback | 34.1 | 15.9 | 4.33 |
| Thumbs problem | 40.3 | 20.3 | 3.75 |
| Education problem | 24.9 | 10.1 | 5.04 |

The overall means (Table 9) for brainstorming, critical problem solving, and individual brainstorming are significantly different from each other at the .01 level by the Newman-Keuls test (Winer, 1962). By the same test individual brainstorming under both feedback and nonfeedback conditions is clearly superior (.01) to any of the other conditions (Table 10). Contrary to the author's prediction, there is no main effect due to feedback. The Thumbs problem elicits more responses than the Education problem. There are no significant interactions. In order to make a more sensitive test of differences between feedback and nonfeedback conditions under the brainstorming and critical problem-solving procedures, the individual brainstorming condition was dropped from the analysis and the ANOVA recalculated. See Table 12.

The main effect for procedure is still significant at the .01 level. A test of simple main effects, which are presented in Table 10, indicates that the difference between the brainstorming groups and critical problem-solving groups under the nonfeedback condition falls short of significance ($p < .10$). As predicted, the difference between the feedback conditions is significant ($p < .01$). The difference, however, is not due to improved performance under the brainstorming conditions, but rather to poor performance under the critical problem-solving conditions. There are no significant differences between the feedback and nonfeedback groups within any of the procedures.

TABLE 10
MEANS FOR SIMPLE MAIN EFFECTS—20 MINUTES OF WORK

| Criterion | Group brainstorming | | Critical problem solving | | Individual brainstorming | |
|----------------------|---------------------|----------|--------------------------|----------|--------------------------|----------|
| | No feedback | Feedback | No feedback | Feedback | No feedback | Feedback |
| Total both problems | | | | | | |
| Total no. ideas | 30.3 (25.5) | 29.3 | 23.7 (19.4) | 18.3 | 48.2 (41.1) | 45.9 |
| No. good ideas | 13.9 (12.2) | 14.3 | 10.9 (9.7) | 8.8 | 22.8 (19.7) | 20.6 |
| Mean quality ideas | 4.14 (4.21) | 4.48 | 4.44 (4.41) | 4.50 | 4.43 (4.49) | 4.37 |
| Thumbs problem | | | | | | |
| Total no. ideas | 37.5 (31.3) | 38.3 | 31.0 (25.3) | 21.4 | 57.3 (49.0) | 55.6 |
| No. good ideas | 18.5 (16.0) | 19.7 | 15.2 (14.0) | 11.3 | 29.7 (25.9) | 27.3 |
| Mean quality ideas | 3.64 (3.68) | 3.82 | 3.63 (3.72) | 3.92 | 3.77 (3.84) | 3.68 |
| Education problem | | | | | | |
| Total no. ideas | 23.1 (19.7) | 20.3 | 16.1 (13.5) | 15.1 | 38.7 (33.2) | 35.9 |
| Total no. good ideas | 9.3 (8.3) | 8.8 | 6.7 (5.3) | 6.3 | 15.8 (13.5) | 13.8 |
| Mean quality ideas | 4.63 (4.75) | 5.14 | 5.24 (5.10) | 5.07 | 5.07 (5.13) | 5.07 |

Note.—Numbers in parentheses are scores for the first 15 min. of work under the No Feedback conditions.

TABLE 11
SUMMARY ANALYSIS OF VARIANCE FOR TOTAL NUMBER OF IDEAS, NUMBER OF GOOD IDEAS,
AND MEAN QUALITY OF IDEAS—20 MINUTES OF WORK

| Source of variation | df | Total no. ideas | | No. good ideas | | Mean quality of ideas | |
|---|----|-----------------|---------|----------------|--------|-----------------------|--------|
| | | MS | F | MS | F | MS | F |
| Between Ss | 35 | | | | | | |
| Procedure (A) | 2 | 4209.29 | 50.59* | 857.04 | 52.74* | .15 | |
| Feedback (B) | 1 | 147.35 | 1.77 | 30.68 | 1.89 | .24 | |
| A × B | 2 | 29.85 | | 12.95 | | .26 | |
| Ss within groups Error _g | 30 | 83.19 | | 16.25 | | .32 | |
| Within Ss | 30 | | | | | | |
| Problems (C) | 1 | 4247.35 | 106.32* | 1850.34 | 79.78* | 30.08 | 64.97* |
| A × C | 2 | 113.44 | 2.84 | 71.85 | 3.09 | .09 | |
| B × C | 1 | 7.35 | | 2.35 | | .01 | |
| A × B × C | 2 | 63.71 | 1.60 | 10.18 | | .25 | |
| C × Ss within groups Error _w | 30 | 39.95 | | 23.23 | | .46 | |

* $p < .01$.

TABLE 12
SUMMARY ANALYSIS OF VARIANCE FOR TOTAL NUMBER
OF IDEAS, NUMBER OF GOOD IDEAS—20 MINUTES
OF WORK, INDIVIDUAL BRAINSTORMING
EXCLUDED

| Source of variation | df | Total no. ideas | | No. good ideas | |
|---------------------------|----|-----------------|--------|----------------|--------|
| | | MS | F | MS | F |
| Between Ss | 23 | | | | |
| Procedure (A) | 1 | 936.34 | 13.06* | 212.52 | 14.50* |
| Feedback (B) | 1 | 120.34 | 1.68 | 9.76 | |
| A × B | 1 | 56.32 | | 16.95 | 1.16 |
| Ss within groups | | | | | |
| Error _b | 20 | 71.68 | | 14.66 | |
| Within Ss | 24 | | | | |
| Problems (C) | 1 | 2160.09 | 54.45* | 841.69 | 34.45* |
| A × C | 1 | 90.24 | 2.27 | 31.69 | 1.30 |
| B × C | 1 | 18.74 | | 1.95 | |
| A × B × C | 1 | 114.10 | 2.88 | 20.59 | |
| C × Ss within | | | | | |
| groups Error _w | 20 | 39.67 | | 24.43 | |

* *p* < .01.

Mean Quality of Ideas

For mean quality of ideas, the only significant main effects are due to differences between problems. The Education problem receives a higher mean rating than the Thumbs problem. There are no simple main effects or interactions.

A second way of looking at these results is to compare the various conditions in terms of amount of time spent working directly on the problems. The feedback group spent their second 5 min. listening to the

TABLE 13
MEANS FOR MAIN EFFECTS—15 MINUTES OF WORK

| Condition | Total no. | No. good | Mean quality |
|--------------------------|-----------|----------|--------------|
| Brainstorming | 27.4 | 13.2 | 4.35 |
| Critical problem solving | 18.9 | 9.3 | 4.44 |
| Individual brainstorming | 43.5 | 20.1 | 4.43 |
| Feedback | 28.7 | 13.8 | 4.37 |
| No feedback | 31.2 | 14.6 | 4.44 |
| Thumbs problem | 36.9 | 19.0 | 3.78 |
| Education problem | 23.0 | 9.4 | 5.04 |

tape and were not allowed to verbalize any responses during this time. Thus they spent only a total of 15 min. on the problem. In order to equate amount of time spent on the problem, the last 5 min. of performance were dropped from all of the nonfeedback groups. The means for these comparisons are given in Tables 10 and 13. The ANOVA is given in Table 14.

Total Number of Ideas and Number of Good Ideas

The main effects for procedures are significant at the .01 level (Table 14). All three procedures are significantly different from each other at the .01 level. The Thumbs

TABLE 14
SUMMARY ANALYSIS OF VARIANCE FOR TOTAL NUMBER OF IDEAS, NUMBER OF GOOD IDEAS,
AND MEAN QUALITY OF IDEAS—15 MINUTES OF WORK

| Source of variation | df | Total no. ideas | | No. good ideas | | Mean quality of ideas | |
|----------------------|----|-----------------|---------|----------------|---------|-----------------------|---------|
| | | MS | F | MS | F | MS | F |
| Between Ss | 35 | | | | | | |
| Procedure (A) | 2 | 3753.18 | 50.41** | 727.10 | 44.44** | .65 | 2.17 |
| Feedback (B) | 1 | 115.02 | 1.54 | 9.49 | | .10 | |
| A × B | 2 | 59.10 | | 12.88 | | .22 | |
| Ss within groups | | | | | | | |
| Error _b | 30 | 74.45 | | 16.36 | | .30 | |
| Within Ss | 30 | | | | | | |
| Problems (C) | 1 | 3486.13 | 99.18** | 168.20 | 77.73** | 28.69 | 58.55** |
| A × C | 2 | 117.55 | 3.34* | 56.29 | | .03 | |
| B × C | 1 | 11.68 | | .23 | | .01 | |
| A × B × C | 2 | 60.00 | 1.71 | 18.51 | | .21 | |
| C × Ss within groups | | | | | | | |
| Error _w | 30 | 35.15 | | 21.64 | | .49 | |

* *p* < .05.
** *p* < .01.

problem elicits significantly more responses than the Education problem. The significant Procedure \times Problem interaction for total number of ideas is due to the fact that the critical problem-solving groups do very poorly on the Thumbs problem relative to the other problem and procedures. Individual brainstorming under both feedback and non-feedback conditions is again clearly superior to any other condition. When the individual brainstorming condition and test for simple main effects are dropped, one finds, as in the previous analysis, that the brainstorming and critical problem-solving procedures do not differ under the nonfeedback conditions.

In order to assess the effects of feedback and procedures more directly, performance under each of the group-problem-solving procedures was divided into 5-min. periods. Both problems were combined in order to increase the stability of the measures. The profiles for total number of ideas and number of good ideas are quite similar so only the latter are presented. The data are shown in Figures 1 and 2. The analysis of variance (Winer, 1962, p. 302) is presented in Table 15. In the contrast for the nonfeedback conditions, the only significant variation between the slopes is in the first 5 min. of work ($p < .05$). The main effect for procedures is not significant as in the previous

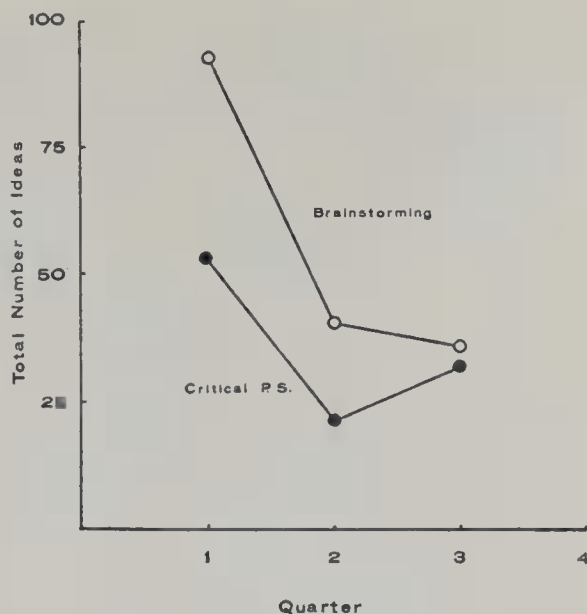


FIG. 2. Number of good ideas produced under the brainstorming and critical problem solving, feedback conditions, across quarters.

analysis. The Procedure \times Quarter interaction is not significant. This indicates that under both procedures Ss run out of ideas at the same rate. When the feedback conditions are contrasted, a significant main effect as in the previous analysis and a significant Procedure \times Quarter interaction are found. A test of simple main effects indicates that the profiles differ significantly at both the first quarter ($p < .01$) and at the second quarter ($p < .05$). It is clear that the significant difference between procedures following feedback is due to its detrimental effect on the critical problem-solving procedure. This conclusion is reinforced by the

TABLE 15

SUMMARY ANALYSIS OF VARIANCE OF NUMBER OF GOOD IDEAS FOR BRAINSTORMING AND CRITICAL PROBLEM SOLVING UNDER BOTH FEEDBACK AND NON-FEEDBACK CONDITIONS OVER QUARTERS

| Source of variation | df | Feedback | | Nonfeedback | |
|-----------------------------|----|----------|--------|-------------|--------|
| | | MS | F | MS | F |
| Between Ss | 11 | | | | |
| Procedure (A) | 1 | 38.02 | 19.47* | 28.52 | 2.71 |
| Ss within groups | 10 | 4.52 | | 10.52 | |
| Within Ss | 36 | | | | |
| Quarter (B) | 3 | 297.97 | 55.90* | 140.19 | 22.18* |
| A \times B | 3 | 27.81 | 5.22* | 8.19 | 1.30 |
| B \times Ss within groups | 30 | 5.33 | | 6.32 | |

* $p < .01$.

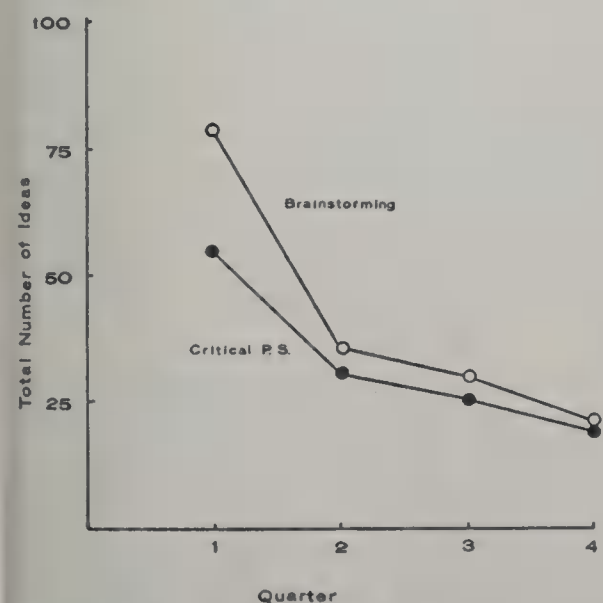


FIG. 1. Number of good ideas produced under the brainstorming and critical problem solving, nonfeedback conditions, across quarters.

fact that the critical groups tend to compensate for this inferior performance by doing somewhat better during the third quarter. Both of these factors are responsible for the significant interaction.

Discussion

Induction of Instructions

One of the main purposes of the experiment was to test the heuristics, described earlier, as adequately as possible, while controlling for important interpersonal factors in small groups. This was attempted primarily by a very careful specification of instructions.

Were the instructional sets adequate? Considering the results in Table 8, it is clear that the critical groups showed somewhat greater cohesiveness, motivation, enjoyment, and perhaps more satisfaction with their procedure than the brainstorming groups. They did not feel they worked in a more critical atmosphere nor did they report feeling more nervous than the brainstorming groups. From a group dynamics standpoint, then, it could not be argued that their performance would be more disrupted by interpersonal factors than it would be in brainstorming groups. The reverse is more likely to be true. Since the instructions were designed to forestall problems in these areas, particularly in the critical problem-solving groups, we feel we were very successful in establishing the atmosphere we had hoped to achieve.

Performance

As predicted, individual brainstorming under both feedback and nonfeedback conditions is by far superior to either group brainstorming or critical problem solving under feedback or nonfeedback conditions. This finding clearly confirms the studies cited earlier. It should be noted, however, that it sheds no light on the possibility that training might improve performance under these conditions, such that group work would be equal or superior to individual work. Second, we may be contrasting groups of a size that maximize the differences between individual and group work. There may be more rapidly

diminishing returns as N increases in nominal groups than in real groups (see also, Thomas & Fink, 1963, and Utterback & Fotheringham, 1958). The data in this experiment suggest that this may very well be the case. Both of these possibilities need to be systematically assessed before any closure on the question of group versus individual work is reached.

The failure to find a significant effect due to feedback either as an interaction or as a main effect runs counter to the expectations of this study. A detailed analysis reveals that, if anything, feedback is a detrimental procedure for critical problem-solving groups. For brainstorming both individual and group, it appears that feedback in the form used here can be as effective as continuous work, but no more so.

The significant overall difference between critical problem solving and brainstorming masks a number of important facts which become apparent when the data are analyzed in detail. The curves in Figures 1 and 2 suggest that the major difference between critical problem solving and brainstorming occurs during the first 5 min., then the curves tend to converge. As noted earlier, number of ideas is a function of the limited number of possible responses. Note, however, that the curves do not cross. In the nonfeedback condition the brainstorming groups do consistently, but not significantly, better over every period. In the feedback condition the critical groups show the negative effects of feedback during the second period. In conjunction with the experiments cited earlier, it seems fair to conclude that brainstorming is superior to critical problem solving both when responses are written and taped, but this superiority is rather small in size.

If we extrapolate to the practical situation where group procedures often must be used, these results have a number of implications. First, if all one wants is a fixed number of good ideas, and this number is not too large, brainstorming is no better than critical problem solving and probably not worth the trouble. Second, if the payoff:cost ratio is small, that is, if just a few more good ideas could be very valuable, it is preferable to use

brainstorming. These conclusions hold for both the equal man-hour and equal work-time comparison. If one takes into account that the brainstorming procedures were novel and somewhat strange to the Ss (none of them had seen or heard more than a peripheral discussion of brainstorming), it may be that these conclusions are too restrictive. That is, training may help brainstorming groups more than critical problem-solving groups. Also, as with the comparison between group and individual brainstorming, these conclusions may have to be seriously modified for large groups. Group size and amount of training are two variables that should be systematically varied in future studies. Proponents of brainstorming have always suggested larger groups than those used in this and previous experiments, and it seems plausible that brainstorming would be more effective than critical problem solving in such groups.

The results of the feedback-nonfeedback manipulation are of some importance in spite of the lack of an overall significant difference or significant interaction. They suggest that, if anything, feedback is a detriment to critical problem-solving groups. For brainstorming, both individual and groups, it appears that feedback in the form used can be as effective as continued work, but not more effective. More important, however, is the fact that real problem-solving groups invariably receive some form of feedback at some time or other. If these results have any generality with respect to various types of feedback (ours could be characterized as both task and socially relevant) and various sources of feedback (e.g., from people other than group members), then brainstorming may be the preferred procedure on this basis also.

PERSONALITY DATA

Prior to participation in the problem-solving part of the experiment, Ss were tested for 1 or 2 hr. They were given the California Psychological Inventory (CPI) (Gough, 1957), the Gough-Sampson College Vocabulary Test (Ss in Experiment I did not take this test) (Gough & Sampson, 1954), and the Firo-B Scales (Schutz, 1958). In Ex-

periment II only some Ss took the Myers-Briggs type indicator (Myers, 1962). The type indicator was administered in an unorthodox fashion. All items not scored on the current scales were crossed out in the test booklet and Ss were told to ignore them. This procedure may have influenced the results in some unknown fashion.

The data from Experiment I consist of the Ss in Groups D, E, and F (see Table 1) plus 11 replications. Five Ss were dropped from the analysis for either failing to take the personality tests or for invalidating them.

Correlations between personality variables and performance, as measured by number of good ideas, are reported in Table 16. The first 18 variables in Table 16 are from the CPI. Variables 19-24 are from the Firo-B, Variables 25-28 are from the Myers-Briggs type indicator, and Variable 29 is the College Vocabulary Test. Variables 30-34 represent five factors extracted from a factor analysis of the CPI.

Factor Analysis

The factor analysis is reported in Table 17 and was obtained from a normal Varimax rotation with 1's in the diagonals (Kaiser, 1959) and included all Ss ($N = 194$). Factor I, Interpersonal Effectiveness, is defined primarily by the first five scales of the CPI, (Do, Cs, Sy, Sp, and Sa). These scales reflect dominance, poise, ascendancy, and self-assurance. This factor corresponds closely to the conventional extroversion factor (Sy) but it also reflects efficient intellectual functioning (see Gough, 1957, p. 36) and high interpersonal effectiveness (Do, Sp, Sa). These two facets of behavior are not generally implied in the concept of extroversion which tends to focus primarily on social and participative aspects of behavior. Factor II, Adjustment, is defined by the Good Impression, Self-Control, Achievement via Conformance, Responsibility, and Sense of Well-Being scales. This factor reflects good adjustment in the form of maturity, social responsibility, and motivation to achieve in conventional ways. Factor III, Intellectual Functioning, is defined by the Flexibility, Achievement via Independence, Tolerance, Psychological-Mindedness, and In-

TABLE 16
CORRELATIONS BETWEEN PERSONALITY VARIABLES AND PERFORMANCE IN VARIOUS GROUP AND INDIVIDUAL PROBLEM-SOLVING CONDITIONS

| Variables | Conditions | | | | | | | | | | | |
|----------------------------------|----------------|------------|----|-------------|--------------------------|----|-------------|----------|--------------------------|-------------|----|----------|
| | Brainstorming | | | | Critical problem solving | | | | Individual brainstorming | | | |
| | Original study | | N | No feedback | Feedback | N | No feedback | Feedback | N | No feedback | N | Feedback |
| | Group | Individual | | | | | | | | | | |
| 1. Dominance | 51 41** | 36** | 24 | 52** | 15 | 20 | 18 | 13 | 21 | 20 | 24 | 12 |
| 2. Capacity for Status | 51 40** | 15 | 24 | 54** | 16 | 23 | 29 | 38* | 24 | 19 | 24 | -10 |
| 3. Sociability | 51 46** | 34* | 24 | 43* | 25 | 23 | 34* | 24* | 24 | 34* | 24 | 15 |
| 4. Social Presence | 51 39** | 17 | 24 | 42* | 51* | 23 | 17 | 17 | 24 | 21 | 24 | -02 |
| 5. Self-Acceptance | 51 51** | 36** | 24 | 44* | 08 | 23 | 32 | 14 | 24 | 17 | 24 | -01 |
| 6. Sense of Well Being | 51 23 | 10 | 24 | 35 | 09 | 20 | -13 | -05 | 21 | 25 | 24 | -29 |
| 7. Responsibility | 51 18 | 02 | 24 | 38 | -20 | 23 | 03 | -10 | 24 | -08 | 24 | -06 |
| 8. Socialization | 51 07 | -05 | 24 | 02 | -10 | 20 | -17 | -01 | 21 | 43* | 24 | -01 |
| 9. Self-Control | 51 -15 | -29* | 24 | -04 | -31 | 23 | -16 | -36 | 24 | -15 | 24 | -01 |
| 10. Tolerance | 51 30* | -07 | 24 | 24 | 13 | 23 | 20 | -11 | 24 | 07 | 24 | 10 |
| 11. Good Impression | 51 02 | -15 | 24 | 25 | -15 | 23 | 20 | -06 | 24 | 12 | 24 | -34 |
| 12. Communality | 51 25 | 28* | 24 | 05 | -01 | 20 | -17 | -16 | 20 | 26 | 24 | 19 |
| 13. Achievement via Conformance | 51 25 | -06 | 24 | -14 | 22 | 23 | 13 | -29 | 24 | 21 | 24 | -08 |
| 14. Achievement via Independence | 51 06 | -25 | 24 | 18 | 08 | 23 | 08 | -13 | 24 | -18 | 24 | -04 |
| 15. Intellectual Efficiency | 51 44** | 19 | 24 | 28 | 19 | 21 | 07 | 25 | 21 | 29 | 24 | 30 |
| 16. Psychological-Mindedness | 51 03 | -14 | 24 | 29 | 16 | 23 | 11 | 01 | 24 | 12 | 24 | 25 |
| 17. Flexibility | 51 10 | -12 | 24 | 24 | 23 | 20 | 07 | -10 | 20 | -11 | 24 | 32 |
| 18. Femininity | 51 -09 | -10 | 24 | -28 | -24 | 23 | -05 | -27 | 24 | -23 | 24 | 38 |
| 19. Expressed Inclusion | 51 40** | 09 | 23 | 42* | -11 | 21 | 44* | 20 | 22 | 16 | 24 | 38 |
| 20. Wanted Inclusion | 51 09 | 07 | 23 | 17 | -08 | 21 | 34 | 18 | 22 | 03 | 24 | -09 |
| 21. Expressed Control | 51 36* | 38** | 23 | 47* | 08 | 21 | 40* | 34 | 22 | -17 | 24 | 11 |
| 22. Wanted Control | 51 10 | -10 | 23 | -69** | 24 | 21 | 08 | -29 | 22 | 07 | 24 | -08 |
| 23. Expressed Affection | 51 13 | -04 | 23 | 43* | 00 | 21 | 51* | -11 | 22 | 29 | 24 | 08 |
| 24. Wanted Affection | 51 09 | -13 | 23 | -22 | -32 | 21 | 50* | -11 | 22 | 15 | 24 | 06 |
| 25. Extroversion-Introversion | 51 -47** | -19 | 15 | -56 | 15 | 13 | -30 | -18 | 9 | -26 | 9 | 18 |
| 26. Sensing-Intuiting | 51 13 | -22 | 15 | 66* | 34 | 13 | -22 | -12 | 9 | 04 | 9 | -20 |
| 27. Thinking-Feeling | 51 04 | 00 | 15 | 16 | 07 | 13 | -48 | -09 | 9 | -36 | 9 | -24 |
| 28. Judging-Perceiving | 51 -28* | -25 | 15 | 29 | 19 | 13 | -34 | 03 | 9 | -46 | 9 | -34 |
| 29. College Vocabulary Test | 51 - | - | 24 | 06 | 38 | 24 | 27 | 09 | 24 | -12 | 24 | -09 |
| 30. Interpersonal-Effectiveness | 51 49** | 37** | 24 | 52** | 22 | 24 | 37 | 37 | 24 | 22 | 24 | 06 |
| 31. Adjustment | 51 -02 | -21 | 24 | 16 | -15 | 24 | 27 | 09 | 24 | -11 | 24 | -27 |
| 32. Intellectual-Functioning | 51 13 | -22 | 24 | 16 | 20 | 24 | -14 | -20 | 24 | 00 | 24 | 39 |
| 33. Conventionality | 51 25 | 20 | 24 | -11 | 05 | 24 | -37 | -35 | 24 | 16 | 24 | 21 |
| 34. Femininity | 51 02 | 00 | 24 | -05 | -29 | 24 | 34 | 00 | 24 | -28 | 24 | 32 |

Note.—A two-tailed test was used in the original study. A one-tailed test was used in all other samples when the variable was significant in the original study, otherwise the test was two-tailed.

TABLE 17

FACTOR ANALYSIS OF CPI SCALES—ORTHOGONAL FACTOR LOADINGS FOR VARIMAX ROTATION

| Variables | Percent communality and loadings for 5 factors | | | | |
|----------------------------------|--|---------------------|--------------------------------------|-------------------------------|--------------------|
| | Interpersonal effectiveness 30.60 | Adjustment 29.49 | Intellectual functioning 17.61 | Conven- tionality 13.46 | Femininity 8.84 |
| 1. Dominance | .83 | .23 | -.08 | .13 | .04 |
| 2. Capacity for Status | .74 | .23 | .30 | -.09 | -.02 |
| 3. Sociability | .89 | .07 | .01 | .15 | -.15 |
| 4. Social Presence | .79 | -.08 | .33 | -.05 | -.28 |
| 5. Self-Acceptance | .88 | -.10 | .04 | .06 | -.02 |
| 6. Sense of Well-Being | .28 | .67 | .27 | .27 | -.28 |
| 7. Responsibility | .06 | .70 | .01 | .32 | .31 |
| 8. Socialization | -.06 | .48 | -.10 | .69 | .04 |
| 9. Self-Control | -.37 | .83 | .20 | .15 | .02 |
| 10. Tolerance | .23 | .56 | .61 | .21 | -.19 |
| 11. Good Impression | .13 | .87 | .09 | -.10 | .01 |
| 12. Communality | .17 | .08 | -.04 | .86 | .00 |
| 13. Achievement via Conformance | .28 | .72 | .06 | .36 | -.01 |
| 14. Achievement via Independence | .06 | .32 | .77 | .10 | .11 |
| 15. Intellectual Efficiency | .52 | .38 | .50 | .24 | -.18 |
| 16. Psychological-Mindedness | .24 | .38 | .50 | -.20 | -.07 |
| 17. Flexibility | .02 | -.25 | .78 | -.30 | .04 |
| 18. Femininity | -.18 | .03 | .03 | -.02 | .92 |

Intellectual Efficiency scales. This factor reflects a flexible, tolerant, independent, and efficient mode of intellectual functioning. Factor IV, Conventionality, is defined by the Communality and Socialization scales. It reflects a tendency to adhere strictly to the rules (Socialization). Factor V, Femininity, is defined exclusively by the Femininity scale. The high end of the Fe scale reflects feminine interests (e.g., appreciative, patient, helpful, sympathetic). For males a high Fe score reflects an openness and willingness to admit unconventional interests and feelings.

In Experiment I the procedure, scoring, and sample of Ss were slightly different from Experiment II. Nevertheless the group brainstorming condition in Experiment I and the group brainstorming nonfeedback condition in Experiment II are similar enough to consider the second a cross-validation of the first.

Individual Brainstorming

The Ss in Experiment I were tested under both group and individual conditions in that order. It was felt that this may have resulted

in a considerable carry-over from the group to the individual situation. Also the individual Ss in Experiment I wrote their responses. If the correlations for the individual condition in Experiment I are compared with the individual brainstorming nonfeedback condition, one sees that there was considerable shrinkage for Dominance, and Self-Acceptance, and a reversal for Expressed Control, the three variables having the highest correlations with the criterion in the original sample. This shrinkage may be due to either (a) a carryover effect in the original sample which did not occur in the replication or (b) the fact that Ss in the replication sample verbalized rather than wrote their responses. In view of the clear-cut cross-validation of the group findings reported below, the former explanation seems to be the most plausible. Under the feedback condition no personality variables are significantly related to performance.

Group Brainstorming

Under the brainstorming nonfeedback condition the first five scales of the CPI cross-validate from the original brainstorming

study. The Tolerance and Intellectual Efficiency scales drop somewhat, but the Expressed Inclusion and Expressed Control scales of the Firo-B, the Extroversion-Introversion scale of the Myers-Briggs type indicator, and Factor I (Interpersonal Effectiveness) of the CPI cross-validate very well. The judging-perceiving scale was not significant in the cross-validation sample, but it did cross-validate, because the magnitude of the correlation was the same in both samples. Two variables which are highly related to performance in the replication sample, but not in the original sample, are the Wanted Control scale of the Firo-B and the Sensation-Intuition scale of the type indicator. Both of these scales had been expected to correlate with performance on the original sample and we were surprised when they did not. The pattern of correlations between the Sensation-Intuition scale and other variables was compared in both samples. The patterns were very similar, indicating that the modification of the testing procedure, or sampling differences, are unlikely to be the reasons for this finding. The Expressed-Affection scale of the Firo-B also relates to performance in the replication sample but not in the original. The College Vocabulary Test is unrelated to performance.

One of the most striking findings of this experiment is that the correlations between personality variables and performance in the brainstorming-feedback groups are very different from those in the brainstorming-non-feedback groups. Only one variable, Social Presence, is significantly related to performance in the brainstorming feedback group. The only other variable with a sizable correlation coefficient in this and all other groups is sociability.

Critical Problem Solving

In the critical problem-solving nonfeedback groups again Sociability is significantly related to performance. The Expressed-Inclusion, Expressed-Affection, and Wanted-Affection scales of the Firo-B also yield significant correlations. Factor I (Interpersonal Effectiveness) and Factor IV (Conventionality) are nearly significant. In the critical

problem-solving feedback groups, only Capacity for Status and Sociability were related to performance. Again Factor I (Interpersonal Effectiveness) and Factor IV (Conventionality) are nearly significant. No variable is significantly related to performance under all five group conditions, but Sociability and Factor I (Interpersonal Effectiveness) are nearly so, and the Firo-B Expressed Control scale is a close third.

Feedback

If only the nonfeedback group problem-solving conditions are considered, Sociability, Expressed Inclusion, and Expressed Control are all significantly related to performance. Given larger samples there is no question that the Myers-Briggs Extroversion-Introversion scale and Factor I (Interpersonal Effectiveness) would also have correlated with performance across all three samples. It should also be noted that the correlations with total number of ideas (not reported here) were similar to those reported here only somewhat lower in magnitude.

No variable is significantly related to performance in both group feedback conditions, but Self-Control and Sociability are nearly so.

In order to test for homogeneity within the samples studied, means on all variables were compared between the three major procedures (brainstorming, critical problem solving, and individual brainstorming) and between feedback and nonfeedback conditions within each procedure. A two-tailed *t* test yielded no significant differences on any variable at the .01 level and 7 significant differences out of 174 at the .05 level. There is no doubt the groups were homogeneous with respect to the variables studied.

DISCUSSION

The results of the analysis of the personality data are very clear. There is no strong and consistent relationship between personality as measured by the CPI, Firo-B, and Myers-Briggs type indicator and all three individual performance conditions in this experiment. There are low correlations across the three individual conditions for Sociabil-

ity, Communality, and Intellectual Efficiency on the CPI and judging-perceiving on the type indicator. Feedback under the individual problem-solving procedure has no effect on overall effectiveness, however, it has a strong interaction with personality type; specifically, it depresses the performance of the type of person who is generally quite effective either alone or in a group. This is indicated by the marked contrast in the correlations between the individual brainstorming nonfeedback, and the individual brainstorming feedback groups for a number of scales (Cs, Sy, Sp, Wb, So, Gi, Fe). This finding is reinforced by a high negative correlation ($-.44$, $-.42$) between sociability and satisfaction with group and self-performance under the feedback conditions, as contrasted to the high positive correlations ($+.42$, $+.42$) between the same variables under the nonfeedback condition.

The correlations between personality variables and performance in the various group conditions are more consistent than the correlations between personality variables and performance under various individual conditions. They are not as consistent from group to group as one might hope, and a variety of qualifications are necessary. The cross validation of the original brainstorming group in the brainstorming nonfeedback condition was very successful as Table 16 indicates. There is an interaction between personality and feedback under both group problem-solving procedures, similar to that under the individual procedure. One reason for this effect may be that under feedback conditions the more active subjects react negatively to their own performance and their activity decreases. In the group situation this perhaps spurs other members to perform, since in most groups a minimum level of activity is almost always maintained. Performance on the part of the active individual does not cease altogether, because he realizes that the other active group members sounded no better than he did, and his status with respect to other group members does not change very much. In the individual situation, the active subject has no one with whom to compare himself, and his negative

evaluation of his performance is likely to cause a greater deterioration in his performance, thereby generating a much lower correlation coefficient. This interpretation is reinforced by the very low ratings by individuals, of the effectiveness of the procedure (Table 8). The relative levels of the correlations for the critical feedback and nonfeedback groups do not differ greatly. In line with the tentative explanation given above, we would argue that in the critical problem-solving nonfeedback condition all Ss are actually getting feedback from each other as they go along. Therefore no differential effect would be expected. There are no personality variables that uniquely characterize the nonfeedback condition as opposed to the feedback condition. This indicates that it is unlikely that the feedback procedure stimulates any particular kind of person to perform more effectively. French (1958) has presented evidence that task-feedback has a tendency to improve the performance of achievement-oriented persons. Those results did not generalize to this study where the achievement scales of the CPI were used as indicators of achievement-orientation.

If we look at all the group problem-solving conditions we find only one scale, Sociability, and one factor, Interpersonal Effectiveness, consistently (but not always significantly) related to performance. The Sociability scale is the defining scale of Factor I and fittingly enough it was empirically developed to predict social participation (Gough, 1952). Thus there is no question that Interpersonal Effectiveness is a powerful predictor of problem-solving effectiveness in small groups. Gough's personological characterization of the high scorer on Sociability is very succinct and worth quoting at this point. He reports that high scorers tend to be seen as "outgoing, enterprising, and ingenious, as being competitive and forward, and as original and fluent in thought" (Gough, 1957, p. 10).

If we restrict ourselves to effective performance in just the group brainstorming nonfeedback conditions a larger number of individual variables come into play. Some of these variables differ from Sociability and

are similar to each other in that they reflect a greater degree of dominance, control, aggressiveness, and self-seeking, (Dominance, Expressed-Control, Capacity for Status). The others reflect self-confidence, spontaneity, and expressiveness (Social Presence, Self-Acceptance). The personological implications of these added variables are quite clear. High scoring subjects in the brainstorming groups have well developed social skills, are outgoing, enterprising, original, verbally fluent, fluent in thought, somewhat aggressive, dominant, and controlling, and yet concerned with feelings of others. They possess self-assurance and are spontaneous, expressive, and enthusiastic.

The failure to find any significant correlations between our adjustment factor (Factor II) and performance under any of the group conditions runs counter to Heslin's (1964) general conclusion that adjustment is fairly consistently related to performance measures in small groups. This discrepancy may be accounted for by the fact that both our criterion and predictor variables differ considerably from those used in the studies he reviewed. Nevertheless our results decrease the generality of his conclusions considerably.

In view of the magnitude of the relationships between personality variables and performance reported here, it seems reasonable to point out that the task was relatively neutral and required a minimal amount of interpersonal interaction. It is not at all unlikely that personality factors may be of even greater importance when the task has a high personal valence, and the participants are required to interact to a greater degree.

The data reported here suggest a number of interesting possibilities for future research. The correlations between personality and performance under the various group problem-solving procedures suggest that proper subject selection can increase performance in brainstorming groups more than in critical problem-solving groups. Studies of homogeneous and heterogeneous groups should focus primarily on measures of interpersonal effectiveness. Earlier, reference was made to the use of complex predictors to predict per-

formance in groups. These data indicate that personality variables would definitely be of value in such an endeavor. The correlations between CPI measures and measures of intellectual functioning (Gough, 1957, p. 36) indicate, nevertheless, that the former are not entirely free of general ability. Our failure to find an effect due to feedback should be viewed with caution since extended practice might show a positive effect, particularly in the brainstorming groups. This possibility should be investigated. Unlike Taylor et al. (1958) we feel that extended practice and perhaps more specific instructions would benefit groups more than individuals. Some of the procedures suggested by Gordon (1961) would be likely candidates. We have also suggested that studies contrasting real groups versus nominal groups should systematically vary group size. During construction of the nominal groups, it was readily apparent to us that, in spite of having contributed a large number of ideas, the fourth individual was contributing little to the nominal group. Thus there may be more rapidly diminishing returns in nominal groups than real groups. The fact that individuals can be selected for group work on the basis of personality, without necessarily selecting against individual ability and perhaps in favor of it, suggests another source of differential improvement for group performance. Therefore, in spite of the negative findings concerning group performance reported in this study, there is little doubt that the question of group versus individual performance needs more extensive investigation.

REFERENCES

- BARRON, F. The psychology of creativity. In T. Newcomb (Ed.), *New directions in psychology*. Vol. 2. New York: Holt, Rinehart & Winston, 1965.
- BOUCHARD, T. J. *Personality, problem-solving procedure and performance in small groups*. (Doctoral dissertation, University of California, Berkeley) Ann Arbor, Mich.: University Microfilms, 1967. No. 67-5007.
- BREER, P. E. Predicting interpersonal behavior from personality and role: Unpublished doctoral dissertation, Harvard University, 1960. (See P. A. Hare, *Handbook of small group research*. Glencoe: Free Press, 1962.)

- BRILHART, J. K., & JOCHEM, L. M. Effects of different patterns on outcomes of problem-solving discussion. *Journal of Applied Psychology*, 1964, **48**, 175-179.
- CAMPBELL, J. P. Individual versus group problem solving in an industrial sample. *Journal of Applied Psychology*, 1968, **52**, 205-210.
- CARTWRIGHT, D., & ZANDER, A. (Eds.) *Group dynamics*. (2nd ed.) Evanston, Ill.: Row, Peterson, 1960.
- CATTELL, R. B., SAUNDERS, D. R., & STICE, G. F. The dimensions of syntality in small groups: The neonate group. *Human Relations*, 1953, **6**, 331-356.
- CATTELL, R. B., & STICE, G. F. Four formulae for selecting leaders on the basis of personality. *Human Relations*, 1954, **7**, 493-507.
- CRONBACH, L. J. The two disciplines of scientific psychology. *American Psychologist*, 1957, **12**, 671-684.
- DUNNETTE, M. D. Are meetings any good for solving problems? *Personnel Administration*, 1964, March-April, 12-29.
- DUNNETTE, M. D., CAMPBELL, J., & JAASTAD, K. The effects of group participation on brainstorming effectiveness for two industrial samples. *Journal of Applied Psychology*, 1963, **47**, 30-37.
- FIEDLER, F., MEUWESE, W., & OONK, S. An exploratory study of group creativity in laboratory tasks. *Acta Psychologica*, 1961, **18**, 100-119.
- FRENCH, E. G. Effects of the interaction of feedback and motivation on task performance. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. New York: Van Nostrand, 1958.
- GHISELLI, E. E., & LODAHL, T. M. Patterns of managerial traits and group effectiveness. *Journal of Abnormal and Social Psychology*, 1958, **57**, 61-66.
- GOLANN, S. E. Psychological study of creativity. *Psychological Bulletin*, 1963, **60**, 548-565.
- GORDON, W. J. J. *Synectics: The development of creative capacity*. New York: Harper, 1961.
- GOUGH, H. G. Predicting social participation. *Journal of Social Psychology*, 1952, **35**, 227-233.
- GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologist Press, 1957.
- GOUGH, H. G., & SAMPSON, H. *The College Vocabulary Test, Forms A and B*. Copyright, 1954. Berkeley, University of California, Institute of Personality Assessment and Research.
- GRACE, H. A. Conformance and performance. *Journal of Social Psychology*, 1954, **40**, 333-335.
- HAYTHORN, W. The influence of individual members on the characteristics of small groups. *Journal of Abnormal and Social Psychology*, 1953, **48**, 276-284.
- HAYTHORN, W., COUCH, A. (S.), HAEFNER, D., LANGHAM, P., & CARTER, L. F. The effects of varying combinations of authoritarian and equalitarian leaders and followers. *Journal of Abnormal and Social Psychology*, 1956, **53**, 210-219.
- HESLIN, R. Predicting group task effectiveness from member characteristics. *Psychological Bulletin*, 1964, **62**, 248-256.
- HOFFMAN, L. R. Homogeneity of member personality and its effect on group problem-solving. *Journal of Abnormal and Social Psychology*, 1959, **58**, 27-32.
- HOFFMAN, L. R. Group problem-solving. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965.
- HOFFMAN, L. R., & CLAGETT, G. S. Some factors affecting the behavior of members of problem-solving groups. *Sociometry*, 1960, **23**, 273-291.
- HOFFMAN, L. R., & MAIER, N. R. F. Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *Journal of Abnormal and Social Psychology*, 1961, **62**, 401-407.
- HOROWITZ, M. W., & NEWMAN, J. B. Spoken and written expression: An experimental analysis. *Journal of Abnormal and Social Psychology*, 1964, **68**, 640-647.
- HYMAN, R. On suspending judgment while producing ideas. Paper read at the meeting of the Eastern Psychological Association, New York City, April 1960.
- JONES, E. E., & THIBAUT, J. W. Interaction goals as bases of inference in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person, perception and interpersonal behavior*. Stanford: Stanford University Press, 1958.
- KAISER, H. F. Computer program for Varimax rotation in factor analysis. *Educational and Psychological Measurement*, 1959, **19**, 413-420.
- KELLEY, H. H., & THIBAUT, J. W. Experimental studies of group problem-solving and process. In G. Lindzey (Ed.), *Handbook of social psychology*. Reading, Mass.: Addison-Wesley, 1954.
- LEAVITT, H. J. Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology*, 1951, **46**, 38-50.
- LORGE, I., FOX, D., DAVITZ, J., & BRENNER, M. A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. *Psychological Bulletin*, 1958, **55**, 337-372.
- MACKINNON, D. W. The nature and nurture of creative talent. *American Psychologist*, 1962, **17**, 484-495.
- MAIER, N. R. F. Assets and liabilities in group problem-solving: The need for an integrative function. *Psychological Review*, 1967, **74**, 239-249.
- MANN, R. D. A review of the relationship between personality and performance in small groups. *Psychological Bulletin*, 1959, **56**, 241-270.
- MCGRATH, J. E., & ALTMAN, I. *Small group research: A synthesis and critique of the field*. New York: Holt, Rinehart & Winston, 1966.
- MEADOW, A., & PARNES, S. J. Evaluation of training in creative problem-solving. *Journal of Applied Psychology*, 1959, **43**, 189-194.

MEADOW, A., PARNES, S. J., & REESE, H. Influence of brainstorming instructions and problem sequence on a creative problem solving test. *Journal of Applied Psychology*, 1959, 43, 413-416.

MYERS, I. B. *Manual: The Myers-Briggs type indicator*. Princeton, N. J.: Educational Testing Service, 1962.

NEWELL, A., SHAW, J. C., & SIMON, H. A. Report on a general problem-solving program. In, *Proceedings of the Conference on Information Processing*. Paris: UNESCO, 1960.

OSBORN, A. F. *Applied imagination*. (Rev. ed.) New York: Scribners, 1957.

PARLOFF, M. B., & HANDLON, J. H. The influence of criticalness on creative problem-solving dyads. *Psychiatry*, 1964, 27, 17-27.

PARNES, S. J., & MEADOW, A. Effects of "brainstorming" instructions on creative problem solving by trained and untrained subjects. *Journal of Educational Psychology*, 1959, 50, 171-176.

PARNES, S. J., & MEADOW, A. Evaluation of persistence of effects produced by a creative problem-solving course. *Psychological Reports*, 1960, 7, 357-361.

PARNES, S. J., & MEADOW, A. Development of individual creative talent. In C. W. Taylor & F. Barron (Eds.), *Scientific creativity: Its recognition and development*. New York: Wiley, 1963.

PELZ, D. C. Some social factors related to performance in a research organization. *Administrative Science Quarterly*, 1956, 1, 310-325.

SALVATORE, V. Z., WILLIS, J. E., & MACKINNON, W. J. Group effectiveness in creative problem-solving tasks: An examination of relevant variables. *Journal of Psychology*, 1966, 62, 111-137.

SCHUTZ, W. C. *Firo: A three dimensional theory of interpersonal behavior*. New York: Rinehart, 1958.

SMELSER, W. Dominance in cooperative problem-solving interactions. Unpublished doctoral dissertation, University of California, Berkeley, 1958.

SPEISMAN, J. C., & MOOS, R. H. Group compatibility and productivity. *Journal of Abnormal and Social Psychology*, 1962, 65, 190-196.

STEIN, M. I., & HEINZE, S. J. *Creativity and the individual*. Glencoe: Free Press, 1960.

TAYLOR, C. W., & BARRON, F. *Scientific creativity: Its recognition and development*. New York: Wiley, 1963.

TAYLOR, D. W., & BLOCK, C. H. Should group or individual work come first on problems requiring creative thinking when equal time is devoted to each. Dept. of Industrial Administration and Dept. of Psychology, Yale, Univ., 1957 (Tech. Rep., No. 2, Contract NONR 609120, N.R. 150-166).

TAYLOR, D. W., BERRY, P. C., & BLOCK, C. H. Does group participation when using brainstorming facilitate or inhibit creative thinking. *Administrative Science Quarterly*, 1958, 3, 23-47.

THOMAS, E. H., & FINK, C. F. Effects of group size. *Psychological Bulletin*, 1963, 60, 371-384.

TRIANDIS, H. C. Cognitive similarity and communication in a dyad. *Human Relations*, 1960, 13, 175-183.

TRIANDIS, H. C., HALL, E. R., & EWEN, R. B. Member heterogeneity and dyadic creativity. *Human Relations*, 1965, 18, 33-55.

TUCKMAN, B. W. Developmental sequence in small groups. *Psychological Bulletin*, 1965, 63, 384-399.

TUCKMAN, J., & LORGE, I. Individual ability as a determinant of group superiority. *Human Relations*, 1962, 15, 45-51.

UTTERBACK, W. E., & FOTHERINGHAM, W. C. Experimental studies of motivated group discussion. *Speech Monograph*, 1958, 25, 268-277.

WEISSKOFF-JOELSON, E., & ELISEO, T. S. An experimental study of the effectiveness of brainstorming. *Journal of Applied Psychology*, 1961, 45, 45-49.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw Hill, 1962.

APPENDIX A

Effectiveness Scale

0 = no conceivable contribution to solution of problem; suggestion impossible of attainment
1 = very little, if any contribution to solution of problem; 2 = probably some contribution to solution of problem; 3 = definite minor contribution to solution of problem; 4 = clearly a major contribution to solution of problem.

Probability Scale

0 = very highly improbable or clearly impossible; 1 = conceivable, but improbable; 2 = possible; 3 = probable; 4 = highly probable.

Practicality-importance Scale

0 = impractical or unimportant; 1 = not too practical or not too important; 2 = somewhat practical or somewhat important; 3 = practical or important; 4 = highly practical or very important.

APPENDIX B

QUESTIONNAIRE

1. How much did you enjoy working with this group of people?

| | | | | | | | | | | |
|-----------------|---|---|---|---|--------------|---|---|---|---|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| | | | | | | | | | | |
| I did not enjoy | | | | | I enjoyed it | | | | | |
| it at all | | | | | very much | | | | | |

2. How would you feel if I told you that your group did not do a good job?

| | | | | | | | | | | |
|-------------------------------------|---|---|---|---|---|------------------------------|---|---|---|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| It would not bother me in the least | | | | | | I would be very disappointed | | | | |

3. How would you feel if I told you that you did not do a good job?

| | | | | | | | | | | |
|-------------------------------------|---|---|---|---|---|------------------------------|---|---|---|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| It would not bother me in the least | | | | | | I would be very disappointed | | | | |

4. How much did you enjoy working on these particular problems?

| | | | | | | | | | | |
|-----------------------------|---|---|---|---|---|--------------------------|---|---|---|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| I did not enjoy them at all | | | | | | I enjoyed them very much | | | | |

5. How satisfied were you with the group's performance?

| | | | | | | | | | | |
|-------------------|---|---|---|---|---|----------------|---|---|---|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Very dissatisfied | | | | | | Very satisfied | | | | |

6. How satisfied were you with your own performance?

| | | | | | | | | | | |
|-------------------|---|---|---|---|---|----------------|---|---|---|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Very dissatisfied | | | | | | Very satisfied | | | | |

7. How did listening to the tape recording of the group's early work affect your later performance?

| | | | | | | | | | | |
|-------------|---|---|---|---------------|---|---|-----------|---|---|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Hindered it | | | | Had no effect | | | Helped it | | | |

8. After listening to the tape how did you feel about suggesting further ideas?

| | | | | | | | | | | |
|--|---|---|---|---|---|--|---|---|---|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Felt less comfortable and was less willing | | | | | | Felt more comfortable and was more willing | | | | |

9. Do you feel that this is an effective way to solve problems?

| | | | | | | | | | | |
|----------------------------|---|---|---|---|---|--------------------------|---|---|---|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Very ineffective procedure | | | | | | Very effective procedure | | | | |

10. How nervous did you feel during the group session?

| | | | | | | | | | | |
|--------------|---|---|---|---|---|--------------------|---|---|---|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Very nervous | | | | | | Not at all nervous | | | | |

11. All groups exert some sort of critical atmosphere. In this group how strongly do you feel you were being judged or criticized by other group members?

| | | | | | | | | | | |
|--------------------------------------|---|---|---|---|---|--|---|---|---|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Felt I was being judged a great deal | | | | | | Did not feel I was being judged at all | | | | |

(Received February 23, 1968)

journal of applied psychology
monograph

Vol. 53, No. 2, Part 2

April 1969

**Instrumentality Theory of Work Motivation:
Some Experimental Results and Suggested Modifications**

By

George Graen

Editor: KENNETH E. CLARK

Published by the American Psychological Association

journal of applied psychology

monograph

Vol. 53, No. 2, Part 2

April 1969

INSTRUMENTALITY THEORY OF WORK MOTIVATION: SOME EXPERIMENTAL RESULTS AND SUGGESTED MODIFICATIONS¹

GEORGE GRAEN²

University of Illinois, Urbana

Instrumentality theory hypothesizes that a person's attitude toward an occurrence (outcome) depends on his perceptions of how that outcome is related (instrumental) to the occurrence of other more or less preferred consequences. In this paper we propose an extension of this theory, describe the results of an experimental design to test deductions flowing from the extended model, and suggest how our results lead to further modifications in the theory. The Ss working in a simulated organization were assigned randomly to the following three treatments: (a) a condition where favorable feedback of high achievement was perceived to be contingent upon effective performance, (b) a condition where Ss received an outcome of money which was not contingent upon effective performance, and (c) a control condition where Ss received neither achievement feedback nor money. Results of this experiment, conducted in a realistic but carefully controlled work setting, show that instrumentality theory predictions of particular levels of job satisfaction and/or job performance are confirmed under only a few rather narrowly specified conditions.

In his book *Work and Motivation* (1964), Fromm reviewed much of the empirical evidence on work motivation within the context of a single theoretical model. His model of work motivation is drawn from "instrumentality" conceptualizations presented earlier by

Peak (1955) and is in turn similar to theories presented by Rotter (1955), Atkinson (1958), and Tolman (1959). Instrumentality theory (Peak, 1955) hypothesizes that a person's attitude toward an outcome (state of nature) depends on his perceptions of relationships (instrumentalities) between that outcome and the attainment of various other consequences toward which he feels differing degrees of liking or disliking (preferences). In essence, then, Peak's theory hypothesizes that a person's attitude toward something (say, racially integrated housing) increases monotonically with the algebraic sum of the products of his perceived instrumentalities between other consequences of integrated housing and his relative preferences for seeing such consequences come about. Some support for this type of linkage has been provided by Peak

¹ This research was supported in part by grants from the National Science Foundation, from the Research Board, University of Illinois, and from the General Electric Foundation. Important contributions were made by a number of the author's colleagues at the University of Minnesota and the University of Illinois, especially the author's mentor, Marvin D. Dunnette. The author would like to thank Merle E. Ace, James H. Davis, Marvin D. Dunnette, Charles L. Hulin, Lloyd G. Humphreys, and Harry C. Triandis for reading earlier versions of this paper.

² Requests for reprints should be sent to the author, Department of Psychology, University of Illinois, Urbana, Illinois 61801.

(1955) and her colleagues (Carlson, 1956; Peak, 1960; Rosenberg, 1956). Vroom (1964) applied this general approach to the area of work motivation by postulating that the valence (preference) for attaining a first-level outcome—such as being in an occupation, on a particular job, or being told that his performance is effective—depends on what he perceives to be the other consequences (second-level outcomes) of that outcome and how attractive these other consequences are. This notion leads directly to statements about the relative level of satisfaction a person may feel toward various first-level circumstances (outcomes) such as certain occupations, jobs, or job accomplishments.

In addition to instrumentality, Vroom introduces expectancy into his model of work motivation. Employing the concept in the same way as Rotter (1955), Vroom defines expectancy as a person's perception (subjective probability) of how his actions may be related to the attainment of first-level outcomes. Thus, for Vroom, the "force" impelling a person to perform a particular job-related action depends on the person's preference (valence) for a first-level outcome and his subjective probability estimate (expectancy) that his action will result in attaining that outcome. The term "force" denotes the relative probability that the action will be emitted. Combining instrumentality with expectancy, hypotheses related to goal or outcome-oriented job behaviors become possible—behaviors such as occupational choice, taking or staying on a particular job, or exerting effort toward becoming effective on a particular job.

With the above statement of Vroom's model in mind, we now propose a modest extension of the model and describe an experiment designed to test certain deductions from the extended model with additional modifications of the model based on results obtained in the experiment.

EXTENSION OF VROOM'S MODEL

This extension of Vroom's formulation draws on a number of additional concepts. First, role concepts are employed to help in defining differences between first-level and second-level outcomes and to help specify

more fully the nature of relationships between various concepts. Second, the model is taken out of the relatively constrained ahistorical approach to a broader historical approach. Third, the model is modified so as not to depend on field theory concepts, which are believed often to contain undesirable surplus meaning.

The present instrumentality theory of work motivation views individual work behavior (both attitudinal and instrumental) as outputs of a work personality-work role system. This theory focuses on a single individual work personality in his work role, and a number of concepts refer to this basic unit. One component of this system, an individual's work personality, is defined, in part, as a person's preferences for various consequences of attaining work roles and his dispositions for perceiving and evaluating various instrumentality and expectancy relationships. The other component of this system, an individual's work role, is defined as a set of behaviors expected by the organization and considered appropriate of an incumbent of a position within the organization. Some examples of work roles are an occupational group member, an incumbent of a particular job, an effective job performer, a leader, and a team member. Work roles must be attained and maintained through performing the expected behaviors in such a way that the resulting performance meets the minimum standards of appropriate behavior. These standards of appropriate behavior imply an evaluation of the expected behaviors by an external agent (e.g., a supervisor) and, also, an organizational contingency between meeting the criteria of appropriate behavior and the attainment or maintenance of that work role. Although one person may attain or maintain several different work roles at one time, this formulation focuses on only one role. For each role, the major variables include the expected behaviors and criteria for appropriate behavior and the contingency between behavior and the attainment or maintenance of that single work role. Moreover, associated with the attainment or maintenance of each work role are role outcomes. Role outcomes are defined as particular outcomes accruing to a person from the attainment or maintenance of work roles. Some examples of

role outcomes are a feeling of achievement, recognition, responsibility, status, money, working conditions, interpersonal relations, and occupational development.

Attraction of Work Roles

The attraction of a work role for an individual depends on the perceived attraction of various role outcomes and the perceived instrumentality of that work role for the attainment of these various role outcomes. The attraction of a role outcome for a person is defined as his preference for attaining that outcome. Attraction is assumed to vary from positive through zero to negative values. As with valence (Vroom, 1964), attraction differs from realized satisfaction: Attraction is viewed as the anticipated satisfaction with an outcome. Although attraction may be viewed by some as synonymous with valence, the present author feels that the concept of valence contains too much "surplus meaning" to be useful, because it seems inexorably embedded in the ahistorical field theory approach.

The second variable determining the attraction of a work role for a person is the perceived instrumentality relationship between the attainment of the work role and the attainment of the various role outcomes. Instrumentality is defined as the degree of belief that the attainment of a particular work role will be followed by the attainment of one or more role outcomes. Instrumentality is viewed as a perceived correlation between the attainment or nonattainment of a particular work role and the attainment or nonattainment of a particular role outcome. Instrumentality relationships are viewed as varying from +1.00 through .00 to -1.00. An instrumentality of +1.00 indicates a belief that the role outcome is certain after the attainment of the work role and impossible without it; an instrumentality of -1.00 indicates the reverse; and an instrumentality of .00 indicates a belief of no relationship between attainment or nonattainment of the work role and of that role outcome. If the chances of receiving a role outcome are improved by the attainment of a work role, this work role has positive instrumentality for the attainment of that role outcome; the work role is viewed as helping to attain that

outcome. In contrast, if the chances of receiving a role outcome are decreased by the attainment of a work role, this work role has negative instrumentality for the attainment of that role outcome; the work role is seen as interfering with the attainment of that outcome.

For all role outcomes, attraction and instrumentality are assumed to combine in a multiplicative manner and summate to produce overall attraction. In this way, both positive and negative attraction and positive and negative instrumentality are allowed to enter the equation. One basic assumption of this model is that people are attracted toward favorable outcomes and away from unfavorable outcomes. This model also reflects the instrumentality assumption that instrumentality moderates the relationships between the attraction of a work role and the attraction of the various role outcomes. The multiplicative manner of combining attraction and instrumentality reflects this latter assumption. It states that positive attraction toward a work role can be enhanced in two different ways: First, role outcomes with positive attraction may combine with positive instrumentality. Second, role outcomes with negative attraction may combine with negative instrumentality. In contrast, positive attraction for a work role may be decreased in two other ways: First, role outcomes with positive attraction may combine with negative instrumentality. Second, role outcomes with negative attraction may combine with positive instrumentality. Moreover, a role outcome can make no contribution to the attraction of a work role if either its attraction or its instrumentality is zero. Thus, if the role outcome had negative attraction (e.g., blame for poor work), the work role would be more attractive to the extent that it interfered with the attainment of this outcome (the degree of negative instrumentality). On the other hand, if the role outcome had positive attraction (e.g., praise for good work), the work role would be more attractive to the extent that it enhanced the attainment of this outcome (the degree of positive instrumentality).

The diagram in Figure 1 represents the extended model of work role attraction. This diagram shows how work role attraction de-

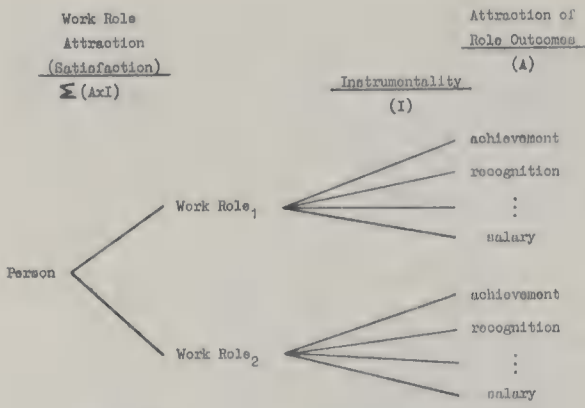


FIG. 1. Extended model of work role attraction.

pend upon the attraction (preference) for various role outcomes and the perceived instrumentality relationship between the work role and the various role outcomes. Moreover, a person's work role satisfaction is predicted from his perceived work role attraction.

Responsivity of Instrumentality

A set of questions that Vroom (1964) does not consider are perceptual consequences of certain work experiences. For example, what are the consequences of attaining a work role and also certain role outcomes on instrumentality or attraction? If instrumentality theory is to possess practical utility, it must be known to what extent perceived instrumentality is influenced by actual work experience as opposed to cognitive manipulations. An experience of attaining a role outcome while in a work role will increase the perceived instrumentality of that work role for attaining like role outcomes only to the extent that instrumentality is influenced by actual work experience. To test this, the following is hypothesized: If a role outcome is attained following the attainment of a work role, higher perceived instrumentality of that work role for the attainment of like role outcomes will result.

Probability of Work Behaviors

The probability that a person will perform an act to attain or maintain a work role depends upon the attraction of that work role and the perceived expectancy that performing the act will lead to the attainment or maintenance of that work role. The attraction of the work role was discussed in the last

section. Now the concept of expectancy must be considered. Perceived expectancy is defined as the degree of belief that an act will lead to the attainment of a work role. Expectancy is viewed as the subjective probability that performing an act will lead to the attainment of a work role. Expectancies are assumed to vary from .00 to 1.00. An expectancy of 1.00 indicates a belief of certainty that performing the act will lead to the attainment of the work role; an expectancy of .00 indicates a belief of certainty that performing an act will not lead to the attainment of the work role; and an expectancy of .50 indicates a belief of complete uncertainty. Although expectancy was described in terms of the attainment of a work role, this same description applies also to the maintenance of a work role.

The attraction of the work role and the expectancy are assumed to combine in a multiplicative manner to produce the probability of the act being performed. In this way, expectancies are allowed to moderate the relationship between the attraction of the work role and the probability of the act. One basic assumption of this model is that behavior is directed toward the attainment of favorable outcomes and away from unfavorable outcomes. A second basic assumption of this model is that the probability of an act depends not only on the attraction of the goal, but also on the relative odds related to striving for the goal. The multiplicative manner of combining the attractiveness of a work role and expectancy reflects this latter assumption. It states that the probability of an act can be enhanced by a high expectancy and decreased by a low expectancy. This postulate thus incorporates the common sense notion that only a Don Quixote would reach for an unreachable star.

The diagram in Figure 2 represents the extended model of the probability of an act being emitted. This diagram shows how work role attraction and expectancy combine to predict the probability of the act.

This extended model differs from Vroom's original model in two major respects. First, this model distinguishes clearly between work roles and role outcomes, whereas Vroom's model at best implies the natures of different

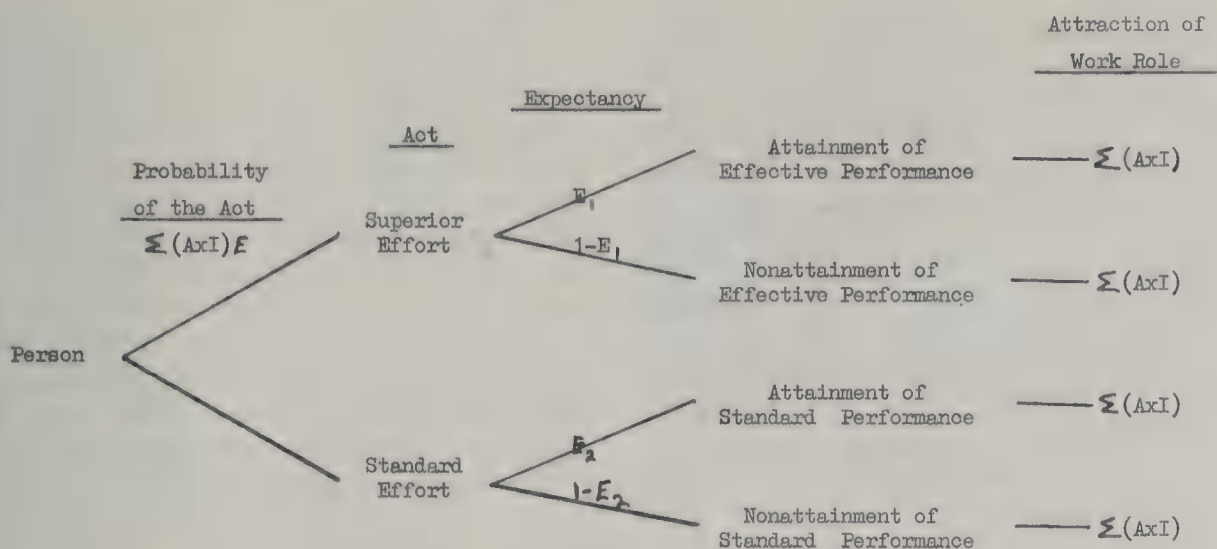


FIG. 2. Extended model of the probability of an act.

kinds of outcomes. In Vroom's original formulation, it often is difficult to determine if his outcomes are more like our work roles than our role outcomes or vice versa. Unless this distinction between kinds of outcomes is made explicitly, predictions from instrumentality theory rapidly degenerate into sterile complexity. If no sharp distinction is made between kinds of outcomes, and thus instrumentality relationships hold between each pair of possible outcomes, the prediction of job satisfaction would be a function of (a) the products of attraction and instrumentalities of the job situation for its "immediate" consequences and (b) the products of attraction and instrumentalities of these "immediate" consequences for "less immediate" consequences.

A second major difference between the two models is the clear developmental implications of the extended model. The emphasis of the extended model on historical as opposed to ahistorical elements allows for better developmental studies of the formulation of work personality and work motivation. This orientation toward understanding the developmental processes of work motivation is the major advantage of the extended model.

EXPERIMENTAL SIMULATION

This developmental model of instrumentality theory hypothesizes causal links between its major variables. These causal hypotheses can be tested only in research de-

signs that contain experimental controls adequate to produce information on the direction of influence between variables. Although increasing the amount of experimental control usually results in increased knowledge as to the direction of influence between variables, it also usually results in a decrease in the scope of the information produced. In terms of research designs, the correlational or field survey design tends to maximize the scope (complexity) of the information produced and minimize the depth (knowledge of directionality) of this information. At the other extreme, laboratory experiments tend to maximize the depth and minimize the scope of the information produced. In contrast to these two extremes, the design used in this study represents an attempt to enhance both the scope and the depth. This type of research design should produce a kind of information not producible by either of the extreme designs.

The design used in this study, called an experimental simulation, attempts to create a situation that contains most of the theoretically important elements of the "real" situation that it is designed to simulate, and at the same time, attempts to maintain as much experimental control as possible. This study was designed to simulate a work organization operating under different organizational climates. Special procedures used to create the simulation included: (a) performing the experiment in a business setting with

many of the usual props (e.g., personnel manager, selection testing, company offices, etc.), (b) providing an organizational purpose for the work, (c) hiring female applicants from the local labor market through a standard recruitment and selection process, (d) using realistic tasks, (e) keeping employees unaware of the fact of the experiment until it had been completed, and (f) embedding the treatment manipulations within an organizational procedure. In contrast to the usual field experiment, in this study experimental control rested fully with the investigator.

The organizational climates that the treatment manipulations were designed to create were (a) a reciprocating climate, (b) a prompting climate, and (c) a control climate. The reciprocating climate is defined as a situation in which the attainment of role outcomes is viewed as contingent upon effective performance. This implies that the organization maintains the practice of reciprocating the effective performance of its members in terms of favorable role outcomes. The prompting climate is a situation in which the attainment of role outcomes is seen as an inducement to effective performance and is not seen as contingent upon effective performance. In the prompting climate, the organization maintains the practice of using role outcomes to stimulate its members toward effective performance without establishing contingencies between effective performance and the subsequent attainment of role outcomes. Finally, the control climate is the situation in which the attainment of role outcomes is viewed neither as being contingent upon effective performance nor as an inducement to effective performance. In the control climate, the organization does not employ role outcomes as motivating conditions. All three of these climates exist within modern organizations.

These three organizational climates were created within the same organization through differential treatment of Ss. For one group of Ss, the organization followed the policy of the reciprocating climate (reward contingent upon effective performance). For a second group, the organization acted in accordance with the policy of the prompting climate (reward as inducement to effective performance). For a third group, the organization practiced

the policy of the control climate (reward neither contingent upon effective performance nor as inducement to effective performance).

The operational manipulations employed to create the organizational climates of reciprocating, prompting, and control were called "achievement feedback," "money," and "control" conditions respectively. Achievement feedback is defined as information from a superior indicating effective performance (outstanding performance) on a previous task. Money is defined as information from a superior indicating an increase in pay directed toward improved performance but not contingent upon previous performance. Finally, control is defined as information from a superior indicating neither effective performance nor a raise in pay directed toward improved performance. Thus, the manipulations differ in terms of the contingencies employed and in terms of the role outcomes used. The nature of the organizational climates of interest require that different role outcomes be employed. The reciprocating climate stipulates that rewards must be contingent upon effective performance. As a consequence of this stipulation, whatever favorable role outcome is placed in the contingency necessarily will also imply recognition for good work and possibly a sense of achievement. In contrast, the prompting climate requires that rewards must not be contingent upon effective performance but merely directed toward it. As a consequence of this requirement, the role outcomes of recognition for good work and a sense of achievement cannot be employed to create the prompting climate.

Within each of the three organizational climates, the study focused on the two work roles of job incumbent and effective performer. Corresponding to these two roles were two prediction models: the job incumbent model and the effective performer model. These two models differ in terms of the kind and number of input variables used to generate predictions and in terms of the variable predicted. The input variables of the job incumbent model are the products of the perceived attraction of each role outcome and the instrumentality of the role of job incumbent for the attainment of like role outcomes—two-term products that are summed alge-

braically over all role outcomes. This job incumbent model predicts overall job satisfaction. In contrast, the input variables of the effective performer model are the products of the perceived attraction of role outcomes and the instrumentality of the role of effective performer for the attainment of like role outcomes that are summed over all role outcomes and then multiplied by the expectancy that increased effort will lead to more effective performance—three-team products. This effective performer model predicts job performance; however, this model also should predict satisfaction to the extent that the effective performance contingencies are important aspects of the overall role.

The specific hypotheses of this study are stated below. The first and third hypotheses are the predicted consequences of job experiences on instrumentality relationships. The second and fourth hypotheses are the predictions of the job incumbent and the effective performer models, respectively.

Hypothesis I. If a role outcome is attained following the attainment of the role of job incumbent, higher perceived instrumentality of that role for the attainment of like outcomes will result.

Hypothesis II. Satisfaction with the role of job incumbent is a monotonically increasing function of the products of the attraction of each role outcome and the perceived instrumentality of that work role for the attainment of like role outcomes summed over all role outcomes (job incumbent model).

Hypothesis III. If a role outcome is attained following the attainment of the role of effective performer, higher perceived instrumentality of that role for the attainment of like outcomes will result.

Hypothesis IV. Job performance is a monotonically increasing function of the product of the attraction of the role of effective performer and the perceived expectancy that increased effort will lead to more effective performance (effective performer model).

DESIGN OF THE EXPERIMENT

Subjects

The Ss in this study were 169 women selected from 203 applicants from the local labor market. Each woman hired was told that the job was a

part-time, temporary position with Decision Systems, Inc., a fictitious company. Most of the Ss were between 15 and 18 yr. old, students in high school, and single. Although the ages ranged from 15 to 66 yr., the sample was most representative of the 15–18 yr. age group.

Instruments

Special instruments developed to measure key parameters from instrumentality theory included: (a) measures of the perceived attraction of various role outcomes (role outcome preferences), (b) measures of the perceived instrumentality of the work roles of a particular job and of effective job performer for the attainment of eight role outcomes, and (c) a measure of the perceived expectancy that increased effort would lead to more effective job performance.

Attraction

Attraction of various role outcomes were measured by an importance questionnaire containing 26 statements about various work role outcomes. Of the 26 statements, 5 each were written for the outcomes of achievement, salary, human relations, and recognition, and 1 each was written for the outcomes of work itself, policies and practices, technical supervision, responsibility, working conditions, and promotion. Instructions asked S to rate each outcome on its importance in a permanent job for him. The response alternatives were "an unnecessary part of the job," "an almost unnecessary part of the job," "an important part of the job," "an almost essential part of the job," and "an essential part of the job." Before rating the outcomes, Ss were instructed to read carefully all 26 items to enable them to evaluate the importance of each statement relative to the other 25 statements. This attraction instrument was pretested and standardized on three different samples: 64 male college sophomores, 77 female college sophomores, and 629 female office workers in one company. The two groups of students were administered the importance questionnaire on two occasions separated by 2 wk. Female office workers were given the questionnaire only once, during working hours at their place of employment.

Instrumentality Measures

Two different sets of instrumentality measures were developed: One set was designed to measure the perceived instrumentality of the work role of a particular job for the attainment of eight different role outcomes, and the other was designed to measure the instrumentality of the work role of an effective job performer for the attainment of the same eight role outcomes. These eight role outcomes were accomplishment, achievement feedback, recognition, responsibility, human relations, company policies and practices, salary, and working conditions. Instructions for the first set of measures asked S to indicate, for each outcome, what he felt were his chances of receiving the outcome on his present job as compared to his previous jobs. The response

alternatives for each outcome were "much worse," "worse," "same," "better," and "much better." These response alternatives were scored 1-5, respectively. Instructions to the second set of measures asked *S* to indicate, for each outcome, what he felt were his chances of receiving the outcome as a result of his effective job performance. The response alternatives were "none," "slight," "fair," "good," and "excellent." These response alternatives also were scored 1-5, respectively.

Expectancy Measure

The particular expectancy relevant to this study was that increased work effort would result in more effective job performance. Instructions for expectancy asked *S* to indicate what he felt were his chances of improving his performance if he "really worked hard." The response alternatives were "No chance at all," "Probably would not improve," "Do not know," "Probably would improve," and "Certain to improve." These alternatives were scored 1-5, respectively.

Treatment Effectiveness Measures

Special measures also were developed to assess the degree to which the treatment manipulation created the intended effects on *Ss*. Two different sets of instruments were developed for this purpose. One set of instruments was contained in a questionnaire designed to measure satisfaction with eight different outcomes of the job. These outcomes were the same ones used in the instrumentality measures, namely, accomplishment, achievement feedback, recognition, responsibility, human relations, policies and practices, salary, and working conditions. In its final version, the satisfaction questionnaire contained scales for the above eight role outcomes and also scales for six outcomes serving as distractors. The purpose of the distractor outcomes was to disguise the outcomes of interest. Each of these 14 scales contained three items. Instructions asked *S* to indicate how satisfied he felt with each outcome of the job described by the item. The response alternatives were "not satisfied," "only slightly satisfied," "satisfied," "very satisfied," and "extremely satisfied." These response alternatives were scored 1-5, respectively. The effectiveness of the treatment manipulations could be checked with these 8 satisfaction scales in the following manner. If achievement feedback had its intended effects, the group receiving this treatment should demonstrate greater satisfaction than the group receiving the control condition with only the outcomes of accomplishment, achievement feedback, and recognition. In contrast, if money had its intended effect, the group receiving this treatment should show greater satisfaction than the control group with only the outcome of salary. Moreover, other differences appearing on these scales would need to be explained.

The second measure of treatment effectiveness was the level of performance *Ss* felt they attained on those tasks completed after receiving the treatment

outcomes. Instructions asked *S* to indicate how he thought he performed compared to his co-workers. Response alternatives were "poorer," "average," "above average," and "among the best." These response alternatives were scored 1-4, respectively. On these measures, if achievement feedback had its intended effect, *Ss* receiving this treatment should indicate a higher level of attained performance than *Ss* receiving the control condition.

Job Satisfaction Measure

Overall job satisfaction was measured by the Hoppock Job Satisfaction Blank (Hoppock, 1935). This instrument contained four items, each with seven response alternatives. Although all four items were asked, only three were scored in the measure of overall job satisfaction. The item dropped asked *Ss* to check the statement best describing how they felt about changing jobs. This item was ambiguous considering the temporary nature of the job.

Work Tasks

The experimental tasks in this study were chosen so that objective performance data could be gathered, so that most *Ss* could learn and master them in a short time, to be sensitive to changes in the expenditure of effort, and to provide what appeared to be "real" work for the "employees." Two tasks meeting all of these requirements are outlined below.

For both work tasks, *Ss* were given a booklet containing 10 pages of computer output for a 155 variable correlation matrix. The output, showing the lower triangle of the matrix, presented one variable at a time with 10 coefficients per row. The coefficients were six-decimal numbers. This booklet was used for two different work tasks. The first task, called the search task, required *Ss* to find certain specified numbers and to write that six-decimal number on an answer sheet. The second task, called the rounding task, required *Ss* to find the specified numbers in the same manner as on the search task, but once the proper number was found, it was to be rounded from six to two decimal places, according to special rules. Only the first two digits of the rounded number were to be written on the answer sheet. Rules for rounding were that *Ss* look at the number in the third place. If the number was 5 or greater, they were to round up one number. If the number was 3 or less, they were to make no change. If the number was 4, *Ss* were to look at the number in the fourth decimal place and apply the rules again. The *Ss* were to continue this process until they either arrived at a number not 4 or exhausted all six decimal places. In the latter case, they were to write down the full number. On both work tasks, *Ss* were to complete each item in the order presented. Performance on each task was measured in two different ways: First, quantity of performance was measured by the number of items attempted. Second, quality of performance was measured by the proportion of items correct of those attempted.

Experimental Instructions

The job was presented as an entirely new kind of computer-related activity. The Ss were told that the usual selection measures did not predict performance on this job. Moreover, Ss were informed that their performance on previous jobs would be a poor indicator of their performance on this job. This rationale was given to justify the blanks that Ss were to complete. In addition, Ss were told that the purpose of this temporary job was to collect cost information to be used in bidding for contract work. The importance of obtaining precise cost estimates for this kind of work was emphasized by outlining the consequences of submitting a bid based upon faulty cost information. The Ss were informed that precise cost information on this work could be gathered only if they worked at a pace that they usually would maintain 8 hr. a day, every working day. This was done in an effort to set the initial amount of effort at a realistic level. Results of a pilot study had shown that Ss maintained unusually high levels of performance when pacing themselves for relatively short work periods. These instructions were designed to increase the pacing period to at least 8 hr. The Ss were told that should the company win the contract, contract work would be performed at a distant city. This was done so that Ss would not expect future employment opportunities with Decision Systems.

Treatments

The achievement feedback, money, and control treatments were presented to Ss individually in a letter from the personnel manager of Decision Systems. Each letter was addressed personally to S and was two pages in length. The first page was the same for all letters, and the second page contained the treatment information. Most of the letter was mimeographed with only the underlined portion shown below being typed. The crucial portion of each treatment letter is shown below. The achievement feedback information was as follows:

According to the results of the work samples which you completed the other day, your performance was AMONG THE BEST of all those who took the samples. *You really did well on this kind of work. If you continue to do as well today, we will let you know. We are extremely happy with your outstanding performance.*

The money information was as follows:

According to the results of the work samples which you completed the other day, your performance was ABOUT AT THE AVERAGE of all those who took the samples. As you know the usual rate of pay for this kind of work is \$1.50 per hour. However, in the hope you will do much better than that today, we will pay you at the rate of \$1.75 per hour.

The control information was as follows:

According to the results of the work samples which you completed the other day, your performance was ABOUT AT THE AVERAGE of

all those who took the samples. As you know the quoted wage for this kind of work is \$1.50 per hour.

CONDUCT OF THE STUDY

The location for the study was a large conference room of a downtown hotel. This room usually was used for conferences and training courses by local business groups. This location was chosen to enhance the simulation of an actual business endeavor.

Data collection was divided into two sessions separated by one day. In the first session, applicants were given a personal history blank and the importance questionnaire, told about the company and the purpose of the work, trained on the tasks, and tested on work samples. Based on this information, Ss were chosen from the applicant pool and assigned to homogeneous ability and outcome preference groups. Within each group, Ss were assigned randomly to the following treatment conditions: (a) achievement feedback, (b) money, and (c) control. In the second session, Ss were given two pretreatment tasks, presented with the treatments, and administered posttreatment measures. Finally, Ss were debriefed and paid. The posttreatment measures included: (a) quantity and quality of performance on four tasks, (b) instrumentality and expectancy measures, (c) the Hoppock measure of overall job satisfaction, and (d) measures designed to assess the effectiveness of the treatment manipulations. These data were analyzed using analysis of variance and correlation analysis.

The study was conducted by E, who was kept entirely unaware of the specific hypotheses under investigation. Moreover, E was not informed of the treatment that any individual S received. This was done to minimize any effects due to E bias (Rosenthal, 1964).

First Session

After the applicants completed the application blank and the importance questionnaire (attraction measure), E read an orientation speech. This orientation speech told the applicants about Decision Systems, Inc., and the particular job. This speech thus contained the overall experimental instructions or set for the entire study. After this experimental set was established, applicants were trained on the search and rounding tasks. Training procedures emphasized a single method of doing the tasks in order to minimize the variance due to different work methods. The applicants who could not learn the tasks during training were rejected from participation in the second session of the study. After applicants were trained, they were given a search task and a rounding task to complete as work samples. Upon completion of the work samples, applicants were allowed to go.

Based on the information obtained in the first session, Ss were chosen from the applicant pool and assigned to homogeneous ability and outcome preference groups. This assignment was performed in the following manner. Scores of applicants on the

attraction of achievement outcomes as opposed to salary outcomes were divided at the median into prefer achievement and prefer salary groups. Ability, as measured by the number of items correct on the two work samples, also was divided at the median into high ability and low ability groups. Each *S* was classified into one of four groups: (a) high ability-prefer achievement, (b) high ability-prefer salary, (c) low ability-prefer achievement, and (d) low ability-prefer salary. Within each of these groups, *Ss* were assigned randomly to one of the three treatment conditions: achievement feedback, money, and control. Finally, within each of these 12 groups, *Ss* were assigned randomly to second sessions. About equal numbers of *Ss* from each of the 12 groups were assigned to each second session. Sampling procedures were successful to a gratifying extent (see Graen, 1967).

Second Session

When *Ss* arrived at the second session, they were given assigned work places. The *E* then read a second orientation speech designed to re-establish the experimental set. Special reference was made to the work samples that *Ss* had completed in the first session, so that when the treatment letters mentioned work samples, *Ss* might make the association readily. Next, training instructions were reviewed and a search and rounding task completed. After these two pretreatment tasks were finished *Ss* were asked to indicate their feelings toward the job. Next, they were given a 10-min. break. During the break, all *Ss* left the work room, and the letters containing the treatments were distributed. After the break, *Ss* were told that they had received letters from the personnel manager of Decision Systems. They were told to read the letters carefully. When *Ss* had finished, the letters were collected. Immediately after this, *Ss* were given four 15-min. tasks: two search and two rounding. Immediately after the last task was completed, *Ss* were given a questionnaire containing the instrumentality, expectancy, overall job satisfaction, treatment effectiveness, and debriefing measures. After this second session, *Ss* were debriefed and paid.

A distinctive feature of this design was that it simulated many important features of the usual employment setting without the loss of experimental control often associated with the conduct of a field experiment (see Weick, 1967). Moreover, treatment and task procedures were refined or modified based on the results of rather extensive pretesting of the instruments and experimental procedures.

Analysis

The sampling design for this study consisted of 12 groups: 3 treatment groups (achievement feedback, money, and control) each containing 4 homogeneous ability and outcome preference groups. Thus, ability and outcome preferences were equated for the 3 treatment groups. In the analysis, the treatment groups are viewed as three samples from different organizational climates and are analyzed

separately. The analysis included the three-way analysis of variance appropriate to this sampling design and correlational analysis (Winer, 1962). The criterion measures of overall job satisfaction and quality and quantity of performance were analyzed both as raw (static) variables and as gain (dynamic) measures. Residual gain scores (Harris, 1963) were employed as measures of the gain from before to after the treatment administration. Residual gain scores are criterion scores with pretreatment scores partialled out (by subtracting from the posttreatment score the score predicted from knowledge of the pretreatment score). Thus, the raw variables are after-only measures, and the residual gain variables are before-after measures. The pretreatment scores used in calculating residual gain scores were those collected during the second session, immediately prior to the administration of the treatments.

RESULTS

Procedural Checks

One set of measures designed to assess the treatment effectiveness of the role outcome manipulations was the set of outcome satisfaction scales. On these measures, if achievement feedback created its intended effects on *Ss*, the group of *Ss* receiving it should show more satisfaction than the group receiving the control outcome with the amount of accomplishment, achievement feedback, and recognition available on the job. As shown in Table 1, the achievement feedback group was more satisfied than the control group with achievement feedback and recognition but not with accomplishment. The nature of the tasks probably accounts for the achievement group not demonstrating higher satisfaction with accomplishment. On the criterion tasks, no *S* completed even half of the available items, and during debriefing, several *Ss* stated a desire to complete the remaining items on the tasks. Apparently, the uncompleted tasks suppressed satisfaction with accomplishment. Turning to the money outcome, if the money treatment had its intended effects, the group receiving it should be more satisfied than the control group with salary, but not with accomplishment, achievement feedback, or recognition. According to Table 1, the money group was more satisfied than the control group on a single scale—salary. In addition to those effects intended for the treatments, no other significant difference was shown on the satisfaction scales.

TABLE 1
MEANS ON THE SATISFACTION SCALES

| Satisfaction with outcomes | Treatment group | | |
|----------------------------|---------------------------------|---------------------------|-----------------------------|
| | Achievement (<i>N</i> = 56) | Money (<i>N</i> = 57) | Control (<i>N</i> = 56) |
| Accomplishment | 9.61 | 9.51 | 9.04 |
| Achievement | | | |
| feedback | 10.43* | 9.51 | 9.46 |
| Recognition | 10.05** | 9.32 | 9.20 |
| Responsibility | 9.79 | 10.07 | 9.55 |
| Human relations | 9.82 | 9.32 | 9.43 |
| Policies & practices | 9.63 | 9.86 | 9.61 |
| Salary | 11.07 | 12.02** | 10.52 |
| Working conditions | 11.75 | 11.42 | 11.02 |

* $p < .05$.

** $p < .01$.

An additional procedural check was the measure of the level of performance Ss felt they had attained on the criterion tasks. If the achievement feedback treatment had its intended effect, Ss receiving it should indicate higher perceived performance than the control group. The money group should not be different from the control group. As shown in Table 2, the perceived level of performance was higher for the achievement group than the control on both the search and the rounding tasks. The money and control groups did not differ. Finally, debriefing results also supported the effectiveness of the treatments.

Hypothesis I

If a role outcome is attained following the attainment of the role of job incumbent, higher perceived instrumentality of that role for the attainment of like outcomes will result.

For this hypothesis to be supported in this

TABLE 2
MEANS ON PERCEIVED LEVEL OF PERFORMANCE
ATTAINED ON THE CRITERION TASKS

| Work task | Treatment group | | |
|-----------|---------------------------------|---------------------------|-----------------------------|
| | Achievement (<i>N</i> = 56) | Money (<i>N</i> = 57) | Control (<i>N</i> = 56) |
| Search | 2.49* | 2.30 | 2.05 |
| Rounding | 2.60* | 2.33 | 2.16 |

* $p < .01$.

TABLE 3
MEANS ON PERCEIVED INSTRUMENTALITY OF THE WORK
ROLE OF JOB INCUMBENT FOR THE ATTAINMENT
OF SELECTED ROLE OUTCOMES

| Instrumentality for outcomes | Treatment group | | |
|------------------------------|---------------------------------|---------------------------|-----------------------------|
| | Achievement (<i>N</i> = 56) | Money (<i>N</i> = 57) | Control (<i>N</i> = 56) |
| Accomplishment | 3.39 | 3.40 | 2.93 |
| Achievement | | | |
| feedback | 3.79** | 3.25 | 3.07 |
| Recognition | 3.68** | 3.04 | 3.05 |
| Responsibility | 3.34 | 3.42 | 3.21 |
| Human relations | 3.34** | 3.02 | 2.96 |
| Policies & practices | 3.48 | 3.40 | 3.32 |
| Salary | 3.95 | 4.19* | 3.80 |
| Working conditions | 3.86 | 3.75 | 3.52 |

* $p < .05$.

** $p < .01$.

study, the achievement feedback group must show higher perceived instrumentality for achievement feedback and recognition outcomes and the money group must show higher instrumentality for salary outcomes. Results of these hypotheses are shown in Table 3. According to Table 3, the achievement feedback group showed higher instrumentality than the control group on achievement feedback, recognition, and human relations. Human relations fits into the achievement feedback, recognition cluster, being concerned with attention to the individual. Moreover, the money group was higher than the control only on salary. These data clearly confirm the hypothesis that the consequence of receiving an outcome following the attainment of the work role of a particular job increases the perceived instrumentality of that work role for the attainment of like outcomes. These data show that instrumentalities can be responsive to actual experience rather than being independent of the external environment. At least when referring to the work role of a particular job, instrumentalities can be enhanced by the actual contingencies between being on the job and receiving various outcomes.

Hypothesis II

Satisfaction with the work role of job incumbent is a monotonically increasing func-

tion of the products of the attraction of each role outcome and the perceived instrumentality of that work role for the attainment of like role outcomes summed over all role outcomes.

After-Only—Job Incumbent Model. Product-moment correlations were calculated between components of the job incumbent model and Hoppock overall job satisfaction to test Hypothesis II in terms of its main predictions and intermediate linkages. This analysis was performed separately for each of the three treatment groups to enable comparisons among the treatments. Results of this analysis are shown in Table 4. In this table, the independent variables of the job incumbent model include the products of the perceived attraction and the perceived instrumentality of the role of job incumbent for attaining: (a) each intrinsic outcome (accomplishment, achievement feedback, recognition, and responsibility), (b) each extrinsic outcome (human relations, policies and practices, salary, and working conditions), (c) all intrinsic outcomes, (d) all extrinsic outcomes, and (e) all outcomes (the complete job incumbent model). Finally, the last independent variable is the sum of only the instrumentalities without attractions.

According to Table 4, the correlations between the complete job incumbent model (the sum of all outcomes) and Hoppock satisfaction supported Hypothesis II for the achievement feedback and the control group, but not for the money group. These correlations were .37, .03, and .35 for the achievement, money, and control groups, respectively. Considering the intrinsic and extrinsic categorization of the role outcomes, the intrinsic class significantly contributed to satisfaction for the achievement and control groups only; the extrinsic class failed to contribute to any of the treatment groups. This difference in strength of relationship for intrinsic and extrinsic classes on satisfaction supports the conclusions from a number of studies (Ewen, Smith, Hulin, & Locke, 1966; Graen, 1966; Graen, in press; Graen & Hulin, in press; Wernimont, 1966) that intrinsic variables are more potent than extrinsic variables in their contributions to overall job satisfaction. Turning to the individual role outcomes, only the intrinsic outcomes were related signifi-

TABLE 4
CORRELATIONS BETWEEN COMPONENTS OF THE JOB INCUMBENT MODEL AND OVERALL JOB SATISFACTION (HOPPOCK)

| Independent variable | Treatment group | | |
|-----------------------------------|-------------------------|-------------------|---------------------|
| | Achievement (N = 56) | Money (N = 57) | Control (N = 56) |
| Attraction X Instrumentality for: | | | |
| Accomplishment | .32** | .04 | .42** |
| Achievement feedback | .20 | -.11 | .26* |
| Recognition | .31* | -.18 | .15 |
| Responsibility | .17 | .08 | .21 |
| Human relations | -.09 | -.05 | .10 |
| Policies & practices | .09 | -.10 | .08 |
| Salary | .20 | .14 | .12 |
| Working conditions | .16 | -.01 | .21 |
| Sum of: | | | |
| Intrinsic outcomes | .36** | -.06 | .36** |
| Extrinsic outcomes | .14 | .00 | .19 |
| All outcomes | .37** | .03 | .35** |
| Instrumentalities only | .62** | .24* | .39** |

* $p < .05$.
** $p < .01$.

cantly to satisfaction: accomplishment for the achievement and control groups, achievement feedback for the control group, and recognition for the achievement group. It should be noted that the most positive observed correlation for the money group was on the outcome of salary. Finally, the sum of the instrumentalities only had significant correlations with satisfaction for all three treatment groups. The magnitude of these correlations were .62, .24, .39 for the achievement, money, and control groups, respectively. As compared to those from the full model, these correlations lend support to the conclusions of Rosenberg (1956) that instrumentalities may be more potent contributions to satisfaction than attractions. It should be emphasized that our money manipulation rendered the group receiving this treatment essentially unpredictable.

After-Only—Effective Performer Model. Although instrumentality theory states no hy-

TABLE 5

CORRELATIONS BETWEEN COMPONENTS OF THE EFFECTIVE PERFORMER MODEL AND OVERALL JOB SATISFACTION (HOPPOCK)

| Independent variable | Treatment group | | |
|--|---------------------------------|---------------------------|-----------------------------|
| | Achievement (<i>N</i> = 56) | Money (<i>N</i> = 57) | Control (<i>N</i> = 56) |
| Expectancy (<i>E</i>) | .43** | .19 | .33** |
| Attraction \times Instrumentality for: Accomplishment <i>X</i> (<i>E</i>) | .43** | .23* | .54** |
| Achievement feedback <i>X</i> (<i>E</i>) | .43** | .04 | .23* |
| Recognition <i>X</i> (<i>E</i>) | .41** | .05 | .34** |
| Responsibility <i>X</i> (<i>E</i>) | .35** | .10 | .35** |
| Human relations <i>X</i> (<i>E</i>) | .15 | .05 | .27* |
| Policies & practices <i>X</i> (<i>E</i>) | .31* | -.01 | .26* |
| Salary <i>X</i> (<i>E</i>) | .26* | .15 | .26* |
| Working conditions <i>X</i> (<i>E</i>) | .30* | -.06 | .28* |
| Sum of: | | | |
| Intrinsic outcomes <i>X</i> (<i>E</i>) | .49** | .14 | .47** |
| Extrinsic outcomes <i>X</i> (<i>E</i>) | .34** | .04 | .32* |
| All outcomes <i>X</i> (<i>E</i>) | .45** | .09 | .42** |
| Instrumentality only <i>X</i> (<i>E</i>) | .61** | .22* | .43* |

* $p < .05$.

** $p < .01$.

hypothesis predicting job satisfaction from the effective performer model, it is reasonable to expect that this work role, being an integral part of the job situation, should be related to overall job satisfaction. The predictions of satisfaction from this model thus were analyzed by correlating the components of the effective performer model and Hoppock satisfaction. Results of this analysis are shown in Table 5. The independent variables in this table include the products of the perceived attraction and the perceived instrumentality of the role of effective performer for attaining the following role outcomes which are then multiplied by expectancy: (a) each intrinsic outcome (accomplishment, achievement feed-

back, recognition, and responsibility), (b) each extrinsic outcome (human relations, policies and practices, salary, and working conditions), (c) all intrinsic outcomes, (d) all extrinsic outcomes, and (e) all outcomes (the complete effective performer model). Finally, the two remaining independent variables are the sum of only the instrumentalities without the attractions and the expectancy that increased effort will lead to more effective performance.

The complete effective performer model was related significantly to satisfaction for the achievement and control groups but not for the money group. These correlations were .45, .09, .42 for the achievement, money, and control groups, respectively. These correlations were as strong as those for the job incumbent model. Again considering the intrinsic and extrinsic categorization, both classes demonstrated significant correlations with satisfaction for the achievement and control groups but not for the money group. This finding was in contrast to that using the job incumbent model. Also in contrast to the results using the job incumbent model, more of the correlations between the individual role outcomes and satisfaction were significant. For the achievement and control groups all correlations involving individual role outcomes were significant, with the exception of human relations outcome for the achievement group. Only accomplishment showed a significant relationship for the money group. As with the job incumbent model, the sum of instrumentalities had significant correlations for all treatment groups: .61, .22, and .43 for the achievement, money, and control groups, respectively. Finally, the correlations between expectancy and satisfaction were .43, .19, and .33 for the achievement, money, and control groups, respectively. The strength of these relationships indicate that the perceived opportunity to influence one's performance is an important determinant of job satisfaction.

Before-After—Job Incumbent. A major difference between after-only and before-after measurement is that the after-only measures are influenced by status differences among Ss, such as differences in abilities, work personalities, and past work experiences, in addition to the effects of the treatments,

whereas the before-after measures minimize the influence of these status differences. These status differences are minimized by using scores that represent the change in the variable of interest in the interval of time between the pretreatment and the posttreatment measures. In this manner, each S serves as his own control and the variation attributable to the influence of the treatment itself is maximized. Thus, gain scores are more appropriate than raw scores for testing the hypotheses of instrumentality theory. In this study, residual gain scores were employed as the measures of change due to their desirable psychometric characteristics (Harris, 1963). The time interval between pre- and post-measures for all gain scores was less than 2 hr.

The correlations of the components of the job incumbent model on the gain in overall job satisfaction are shown in Table 6. The

independent variables are the same as in the analysis of raw satisfaction. According to Table 6, the correlations using the job incumbent model are quite different on the gain in satisfaction from those on raw satisfaction. The predictions of the complete job incumbent model (the sum of all outcomes) was supported only for the achievement group. This was also the case for the intrinsic class of variables, whereas, the extrinsic class showed no significant correlations. Of the individual role outcomes, only achievement feedback and responsibility had significant correlations, and these were only for the achievement group. Finally, the sum of the instrumentalities demonstrated a significant correlation for only the achievement group. These results imply that the generally more positive results employing the raw satisfaction scores probably were reflecting the influences of response sets (Guilford, 1954). These results on the gain in satisfaction support Hypothesis II for the achievement group, but not for the money or control groups.

Before-After—Effective Performer Model. The results of the correlations between components of the effective performer model and gain in overall job satisfaction are shown in Table 7. The independent variables in Table 7 are the same as in the analysis of raw satisfaction using this model. According to Table 7, the correlations between expectancy and gain in satisfaction are .32, .02, and .15 for the achievement feedback, money, and control groups, respectively. As with the job incumbent model, the complete effective performer model (the sum of all outcomes) significantly predicted the gain in satisfaction for only the achievement group. This also was the case for both the intrinsic and the extrinsic classes of role outcomes. In contrast to the results with the job incumbent model all of the correlations of the individual role outcomes were significant for the achievement group, with the exceptions of human relations and working conditions. It should be noted that the outcome of policies and practices was negatively correlated with satisfaction for the money group. Again in contrast to the results using the job incumbent model, the sum of instrumentalities was related significantly for both the achievement and control groups. Finally,

TABLE 6
CORRELATIONS BETWEEN COMPONENTS OF THE JOB INCUMBENT MODEL AND GAIN IN SATISFACTION

| Independent variable | Treatment group | | |
|-----------------------------------|-------------------------|-------------------|---------------------|
| | Achievement (N = 56) | Money (N = 57) | Control (N = 56) |
| Attraction X Instrumentality for: | | | |
| Accomplishment | .19 | -.12 | .10 |
| Achievement feedback | .27* | -.08 | -.06 |
| Recognition | .19 | -.06 | .10 |
| Responsibility | .23* | .15 | .07 |
| Human relations | .02 | -.03 | .05 |
| Policies & practices | .16 | -.19 | -.15 |
| Salary | .14 | .02 | -.13 |
| Working conditions | .18 | .00 | -.01 |
| Sum of: | | | |
| Intrinsic outcomes | .32* | -.04 | .07 |
| Extrinsic outcomes | .19 | -.07 | -.10 |
| All outcomes | .29* | -.07 | -.02 |
| Instrumentalities only | .44** | .05 | .11 |

Note.—Gain in satisfaction scores (residual gains) are the posttreatment Hoppock scores with immediate pretreatment satisfaction partialled out.
* $p < .05$.
** $p < .01$.

TABLE 7

CORRELATIONS BETWEEN COMPONENTS OF THE
EFFECTIVE PERFORMER MODEL AND
GAIN IN SATISFACTION

| Independent variable | Treatment group | | |
|---|---------------------------------|---------------------------|-----------------------------|
| | Achievement (<i>N</i> = 56) | Money (<i>N</i> = 57) | Control (<i>N</i> = 56) |
| Expectancy (E) | .32** | .02 | .15 |
| Attraction \times Instrumentality for: Accomplishment \times (E) | .39** | .06 | .19 |
| Achievement feedback \times (E) | .34** | .00 | .06 |
| Recognition \times (E) | .33** | -.05 | .20 |
| Responsibility \times (E) | .39** | .18 | .17 |
| Human relations \times (E) | .12 | -.04 | .07 |
| Policies & Practices \times (E) | .30* | -.25* | .00 |
| Salary \times (E) | .31* | -.02 | .02 |
| Working conditions \times (E) | .18 | -.06 | .06 |
| Sum of: | | | |
| Intrinsic outcomes \times (E) | .44** | .08 | .20 |
| Extrinsic outcomes \times (E) | .30* | -.12 | .06 |
| All outcomes \times (E) | .40** | -.03 | .13 |
| Instrumentalities only \times (E) | .46** | .02 | .23* |

Note.—Gain in satisfaction (residual gains) are the post-treatment Hoppock scores with immediate pretreatment satisfaction partialled out.

* $p < .05$.

** $p < .01$.

the remarks made in connection with the differences between the results of the after-only and the before-after analyses employing the job incumbent model also apply to these differences using the effective performer model. Moreover, the effective performer model successfully predicted the gain in satisfaction for only the achievement group. When the variance attributable to the influence of the treatments is maximized, neither of the two prediction models of instrumentality theory predicted overall job satisfaction for either the money or the control group.

At this point we will consider the question of the differences between the two work roles

of job incumbent and effective performer. Our results thus far indicate that the predictions from models based upon these two work roles lead to somewhat different predictions. It now would be informative to consider the correlations of the major variables employed in both models. The results of this analysis are shown in Table 8. The correlations in Table 8 are those between corresponding components of the job incumbent model and the effective performer model. The correlation between the complete models (the sum of all outcomes) for the roles of job incumbent and effective performer were .73, .46, and .68 for the achievement, money, and control groups, respectively. These correlations indicate that these two models probably are not tapping very different sources of variation for, at least, the achievement and the control groups. This also is the case for the two classes of outcomes (intrinsic and extrinsic). In contrast, the sum of instrumentalities appears to be measuring similar sources of variation for primarily the achievement group. Although both models deserve further research, the effective performer model should be given the higher priority.

Hypothesis III

If a role outcome is attained following the attainment of the role of effective performer, higher perceived instrumentality of that role for the attainment of like outcomes will result.

If this hypothesis is to be supported in the

TABLE 8

CORRELATIONS BETWEEN COMPONENTS OF THE
JOB INCUMBENT AND EFFECTIVE
PERFORMER MODELS

| Component | Treatment group | | |
|------------------------|---------------------------------|---------------------------|-----------------------------|
| | Achievement (<i>N</i> = 56) | Money (<i>N</i> = 57) | Control (<i>N</i> = 56) |
| Sum of: | | | |
| Intrinsic outcomes | .77 | .36 | .62 |
| Extrinsic outcomes | .64 | .56 | .73 |
| All outcomes | .73 | .46 | .68 |
| Instrumentalities only | .63 | .31 | .47 |

Note.—Components of the effective performer model have been multiplied by expectancy.

TABLE 9

MEANS ON PERCEIVED INSTRUMENTALITY OF THE WORK
ROLE OF EFFECTIVE PERFORMER FOR THE
ATTAINMENT OF SELECTED ROLE
OUTCOMES AND EXPECTANCY

| Instrumentality for outcomes | Treatment group | | |
|---------------------------------|------------------------------|-------------------|---------------------|
| | Achieve- ment (N = 56) | Money (N = 57) | Control (N = 56) |
| Accomplishment | 3.66 | 3.47 | 3.36 |
| Achievement feedback | 3.59* | 3.44 | 3.18 |
| Recognition | 3.46* | 3.25 | 3.02 |
| Responsibility | 3.43 | 3.42 | 3.21 |
| Human relations | 3.11 | 3.11 | 3.13 |
| Policies & practices | 3.52* | 3.37 | 3.21 |
| Salary | 3.52 | 3.68 | 3.39 |
| Working conditions | 3.70* | 3.47 | 3.23 |
| Expectancy | 4.18 | 4.12 | 4.09 |

* $p < .05$.

present study, the achievement feedback group must demonstrate higher instrumentality than the control group between the role of effective performer and the outcomes of achievement feedback and recognition. In contrast, the money group must not show higher instrumentality than the control group for salary outcomes, because the raise in pay was not contingent upon effective performance. This difference is predicted from the nature of the contingencies contained in the achievement feedback and money treatments. Achievement was contingent upon effective performance and money was not.

Results on this hypothesis are shown in Table 9. The achievement feedback group was higher than the control group on achievement feedback, recognition, policies and practices, and working conditions. Differences on policies and practices and working conditions were not predicted from the hypothesis. Possibly, the contingency between effective performance and treatment outcomes generalized to these outcomes. As predicted, perceived instrumentality for the money group was not higher than the control group on salary. Although the treatment group means on salary were in the right direction, the differences were too small to be reliable. These data support the hypothesis that the consequence of

receiving an outcome contingent upon the work role of effective performer is to increase the perceived instrumentality between that work role and like outcomes. These data show that instrumentalities are responsive to actual contingencies rather than being independent of the job situation.

Hypothesis IV

Job performance is a monotonically increasing function of the product of the attraction of the work role of effective job performer and the perceived expectancy that increased effort will lead to effective performance.

After-Only—Effective Performer Model. The task performance measures were quality and quantity scores on two search tasks and two rounding tasks. The quality measure was the number of items correct divided by the number of items attempted, and the quantity score was the number of items attempted. The analysis was the same as that on satisfaction. The results of this analysis were that none of the component variables of the effective performer model demonstrated any significant correlations with any of the task performance measures. Hypothesis IV of instrumentality theory received absolutely no support from this analysis.

Before-After—Effective Performer Model. In this analysis, the influence of status differences among Ss in ability and pretreatment motivation were minimized, and the variance attributable to the influence of the treatments maximized rendering this analysis more appropriate than the raw score analysis. The task measures were residual gain quality and quantity scores on two search and two rounding tasks. Posttreatment tasks were completed immediately after the treatments were administered. The sequence of the posttreatment work tasks was search (E), rounding (F), search (G), and rounding (H). Pretreatment tasks (a search and a rounding task) were completed immediately before the treatment administration. The analysis of performance gain scores was the same as that for satisfaction using the effective performer model. The results of this analysis were that the model predicted consistently across at least two tasks on only the quantity measure of the rounding task—the more complex task. Results on this

ounding task are shown in Table 10. According to Table 10, expectancy correlated with gain in performance .33, $-.23$, .07 on Task F and .46, $-.12$, and .13 on Task H for the achievement feedback, money, and control groups, respectively. These relationships indicate that, in the achievement group, the easier Ss felt it was to improve their performance, the more they improved their performance. In contrast, in the money group, the more difficult Ss felt it was to improve their performance, the more they did improve their performance. The complete model (the sum of all outcomes) successfully predicted the gain in performance on the two rounding tasks for only the achievement group. Further, both the intrinsic and extrinsic classes contributed to the gain in performance for the achievement group. Considering the individual role outcomes, accomplishment, responsibility, and working conditions consistently contributed to the gain in performance for only the achievement group and achievement feedback and salary contributed to the gain in performance

on only a single task. Finally, the sum of the instrumentalities also was related consistently for the achievement group. Considering the full model, the effective performer model did predict consistently for the achievement group and not for the money or control groups. These results on the rounding task support Hypothesis IV of instrumentality theory for only the achievement feedback group and not for the money or control groups.

A partial explanation for the finding that the effective performer model predicted consistently across only the quantity measures of the two rounding tasks is the different reliabilities of the residual gain performance measures. The stability reliability coefficients for the four residual gain measures for each of the treatment groups are shown in Table 11. According to Table 11, the quality scores are less reliable than the quantity scores. Of most interest, the most reliable measure for the achievement feedback group was the quantity measure of the rounding task. If the effective performer model were valid under

TABLE 10

CORRELATIONS BETWEEN COMPONENTS OF THE EFFECTIVE PERFORMER AND THE GAIN IN PERFORMANCE

| Independent variable | Work task | | | | | |
|-----------------------------------|------------|----------|--------|------------|--------|--------|
| | Rounding F | | | Rounding H | | |
| | A | M | C | A | M | C |
| Expectancy (E) | .33** | $-.23^*$ | .07 | .46** | $-.12$ | .13 |
| Attraction X Instrumentality for: | | | | | | |
| Accomplishment X (E) | .28* | $-.01$ | .10 | .44** | .18 | .15 |
| Achievement feedback X (E) | .11 | $-.09$ | .20 | .31* | .03 | .09 |
| Recognition X (E) | .17 | $-.08$ | .02 | .20 | .15 | $-.06$ |
| Responsibility X (E) | .40** | $-.15$ | .10 | .40** | $-.02$ | .00 |
| Human relations X (E) | .16 | $-.09$ | $-.07$ | .19 | $-.07$ | $-.06$ |
| Policies & practices X (E) | .06 | $-.16$ | $-.08$ | .18 | .08 | .00 |
| Salary X (E) | .15 | $-.08$ | .14 | .27* | .03 | .12 |
| Working conditions X (E) | .29* | $-.16$ | $-.14$ | .28* | .02 | $-.07$ |
| Sum of: | | | | | | |
| Intrinsic outcomes X (E) | .29* | $-.12$ | .13 | .41** | .08 | .07 |
| Extrinsic outcomes X (E) | .22* | $-.16$ | $-.05$ | .31* | .03 | .00 |
| All outcomes X (E) | .28* | $-.15$ | .04 | .39** | .06 | .03 |
| Instrumentalities only X (E) | .35** | $-.12$ | .08 | .43** | .05 | .04 |

Note.—Letters A, M, and C represent achievement, money, and control groups, respectively. Gain in performance scores (residual gains) are the posttreatment performance scores with immediate pretreatment performance partialled out.

* $p < .05$.

** $p < .01$.

TABLE 11
STABILITY RELIABILITY ESTIMATES FOR GAIN
IN PERFORMANCE MEASURES

| Measure | Reliability | | |
|-------------------|-------------|-----|-----|
| | A | M | C |
| Search quantity | .53 | .80 | .76 |
| Search quality | .65 | .53 | .12 |
| Rounding quantity | .75 | .71 | .49 |
| Rounding quality | .33 | .16 | .58 |

Note.—Letters A, M, and C represent the achievement, money, and control groups, respectively. Coefficients were calculated by correlating the two posttreatment parallel tasks.

only those conditions that prevailed for the achievement group, given these estimates of reliability, consistent results might be expected on only the quantity measure of the rounding tasks.

Summary of Results

The results of this may be summarized in the following manner:

1. The procedural checks indicated that Ss in each of the conditions were responsive to the appropriate manipulations. The achievement feedback group showed higher satisfaction than the control group with the role outcomes achievement feedback and recognition, and the money group showed higher satisfaction than the control group with the role outcome of salary. Moreover, the achievement feedback group indicated higher perceived performance than the control group on the criterion tasks.

2. The results clearly confirmed the hypothesis (Hypothesis I) that the consequence of receiving a role outcome following the attainment of the role of job incumbent increases the perceived instrumentality of that role for the attainment of like outcomes. The achievement feedback group showed higher instrumentalities than the control group between the role of job incumbent and the role outcomes of achievement feedback, recognition, and human relations. In addition, the money group was higher than the control only on the role outcome of salary.

3. The results concerning Hypothesis II (job incumbent model), considering both the static (raw score) and dynamic (gain score)

analyses, support this prediction of overall satisfaction only for the group performing in the reciprocating climate (achievement feedback group) and not for the groups performing in either the prompting climate (money group) or in the control climate (control group).

4. Employing the effective performer model to predict overall satisfaction, again considering both static and dynamic analyses, resulted in significant correlations for primarily the group performing in the reciprocating climate (achievement feedback group) and possibly in the control climate (control group) but not in the prompting climate (money group).

5. The results regarding Hypothesis III support the prediction that the consequence of receiving a role outcome contingent upon the role of effective performer is to increase the perceived instrumentality between that role and like role outcomes. These data show that instrumentalities were responsive to actual contingencies and were not independent of the organizational climate.

6. The data relevant to Hypothesis IV (effective performer model) predicting job performance support this hypothesis only in the dynamic (gain) analysis and only the group performing in the reciprocating climate (achievement feedback group).

7. In the satisfaction analyses, the differences in results between static (raw score) and dynamic (gain score) analyses probably reflect the influences of response sets and biases.

DISCUSSION

These results taken as a whole indicate that instrumentality theory shows promise of being a scientifically useful model in our attempt to understand work motivation. Employing specially designed measures of the major parameters of instrumentality theory in a simulated organization, this study demonstrated that the job incumbent model can predict job satisfaction and that the effective performer model can predict both job satisfaction and job performance under certain conditions. In addition, this study demonstrated the predicted consequences of certain job experiences on perceived instrumentalities.

The finding that both the job incumbent model and the effective performer model predicted satisfaction and the latter predicted performance under the conditions that existed for the achievement feedback group (reciprocating climate) indicates the presence of certain boundary conditions. What were the important differences between the achievement feedback treatment and the other two treatments? Only in the achievement feedback treatment was the contingency between effective performance and the attainment of a favorable role outcome established in a concrete manner—presenting the achievement feedback contingent upon effective performance. In the control treatment (control climate), this contingency was implied at best but not demonstrated—evaluating previous performance but not presenting a favorable role outcome. In the money treatment (prompting climate), this contingency was undermined by presenting a raise in pay not contingent upon previous performance.

This interpretation was supported by the results of the before-after analysis on overall job satisfaction and on the quantity of performance on the rounding task. The correlations between the expectancy that increased effort would lead to more effective performance and both satisfaction and performance reflected the strength of the established contingencies between effective performance and role outcomes. These correlations on satisfaction were .32, .02, and .15 and on performance were .33, -.23, and .07 for Task F, and .46, -.12, and .13 for Task H for the achievement feedback, money, and control groups, respectively. Therefore, an important boundary condition for instrumentality theory is that contingencies must be established in a concrete manner between effective job performance and attaining favorable role outcomes.

The discovery of boundary conditions for instrumentality theory suggests that one reason other theories of work motivation have not been supported by empirical studies is that they have failed to specify boundary conditions. If the present study had not employed the reciprocating climate, the results of this study would not have supported either model from instrumentality theory. In fact,

the results would have been considered damaging to instrumentality theory. The point of this is that unless the boundary conditions of a theory can be specified, the theory applies under all conditions and can be tested legitimately under all conditions. It is unreasonable to assume that theories of work motivation can be applied to all conditions. An unfortunate consequence of not being able to specify boundary conditions for our theories is that only "wide-band" theoretical formulations incapable of empirical disproof survive to haunt our textbooks and our students.

The discovery of boundary conditions for instrumentality theory, if confirmed, has implications for the design of work organizations. If a goal of a work organization is to understand and predict work role satisfaction and performance, the boundary conditions of instrumentality theory must be designed into the work situation. Instrumentality theory or any theory of work motivation can help to make work behavior understandable only after the boundary conditions have been met. In short, if the goal is to have employees respond to an organization in an understandable and predictable manner, the organization must be designed in such a way that employees perceive it as an understandable and predictable system. Employees' work motivation will be puzzling and unpredictable to the extent that the organization's behavior toward its employees is perceived by the employees as puzzling and unpredictable.

In addition to understanding and prediction, instrumentality theory promises to be useful for the enhancement of work motivation. In this study, it was shown that the perception of instrumentality relationships was responsive to the actual contingencies of the job situation rather than independent of the job environment. Thus, the cognitive manipulations that prove so troublesome for inequity theory (Adams, 1963) were not found in the perceptions of instrumentality relationships. If these results are confirmed, instrumentalities could be enhanced by designing the work situation to produce stronger contingencies between work roles and role outcomes. If this model is valid and the boundary conditions are met, strengthening these contingencies should result in increased satis-

faction and work motivation. Thus, these results suggest the possibility that if work organizations can be designed or restructured to be responsive to the work personalities of individuals, employees' responses to work organizations may be understandable, predictable, and reciprocal.

Instrumentality theory did not successfully predict raw job performance for any of our three treatment groups. In fact, none of the parameters of instrumentality theory showed any consistent relationship to job performance. In addition, Galbraith and Cummings (1967) attempted to test this hypothesis predicting job performance for operatives in a heavy equipment manufacturing company. These authors found so few statistically significant differences relative to the number of their tests that the most reasonable interpretation is that observed differences were due to the operation of chance alone. However, it should not be expected that instrumentality predicts a performance measure that is controlled primarily by status variables. More appropriate tests of the hypotheses derived from instrumentality theory were those on the gains in performance; the effective performer model did predict this criterion for the achievement feedback group. The theory will be considered again with suggested modifications based upon the results of this study.

SUGGESTED MODIFICATIONS

A basic modification of the models of instrumentality theory must be the inclusion of the boundary conditions uncovered in this study. A statement of these hypothesized boundary conditions follows. If the job situation is designed in such a way that performance is evaluated and rewarded with favorable role outcomes and this contingency between effective performance and role outcomes is communicated to employees in a concrete manner, the models of instrumentality theory apply. If these boundary conditions are not met, the models of instrumentality theory can make no predictions. If these boundary conditions are met, the job incumbent model states that job satisfaction is a monotonically increasing function of the algebraic sum of the products of the perceived attraction of various role outcomes and the perceived instrumen-

talities of the work role of being a job incumbent for the attainment of these various role outcomes. In contrast, the effective performer model should be modified drastically to improve its prediction of job performance and satisfaction.

The nature of these needed modifications of instrumentality theory is suggested by the work of Fishbein and his colleagues (Anderson & Fishbein, 1965; Fishbein, 1967) on attitude research. These investigators started with instrumentality theory (Peak, 1955; Rosenberg, 1956) and developed a more general theory of social attitudes. Although Fishbein and his associates have been able to predict satisfaction toward an object using other cognitions about that object, Fishbein (1967) states that knowledge of a person's attitude toward an object does not allow the prediction of the way he will behave toward that object. He also rules out such variables as beliefs about the outcome and behavior intentions toward the outcome. Instead, he proposes to predict behavior by employing a theory of behavior prediction based on Dulany's (1961) theory of propositional control. Fishbein's theory hypothesizes that the probability that an individual will emit a given act, with respect to a given outcome, in a given situation, is a function of the following: (a) his beliefs concerning the consequences of the particular behavior; (b) the attraction of these consequences for him; (c) his beliefs about what he should do under the circumstances; (d) his motivation to comply. Further, the theory specifies that the first two and the last two terms be multiplied and the resulting two products be weighted in a linear regression equation to predict the probability of behavior. These regression weights would differ for each act and each situation.

A theoretical framework—"interdependent role systems"—may be developed by integrating the formulation of Fishbein, several ideas from Katz and Kahn, and the results of this study into the systems approach of modified instrumentality theory.

Overview

In general orientation, interdependent role systems theory views work organizations as

complex sets of interdependent work roles. Each position within an organization may contain a number of these work roles and each role may be interrelated with a number of roles assumed under different positions in the organization. Thus, a position may be related to one set of positions by one work role and related to an entirely different set of positions by a second work role. Work roles are interdependent in the sense that the appropriate behavior specified by one role depends upon the antecedent behavior specified by another's role and/or is required before a subsequent role behavior of another can be performed. In general, work roles are interdependent to the extent that individuals in various related positions have a vested interest in the performance of one another's role behaviors. Work roles are defined as sets of behaviors expected and considered appropriate for an incumbent of the position within the organization. Some requirements of the work role are specified formally by the organization (e.g., stated in terms of operating procedures) while others are determined by the actions of interested colleagues.

Central Role

The key work role for this theory is that of effective performer. The behavior expected and considered appropriate for an individual in this role is that his performance reflects that he is developing and growing as an asset to both his occupation and his organization. An individual may attain and maintain this role without necessarily changing his performance standing relative to others in his work group (in the short run). In this sense, the theory is a model of growth and development rather than a model of performance standing. The details and criteria of the expected behaviors for this role depend on the nature and interrelations of the roles assumed under the position in question and on the nature and interrelations of other interdependent roles. In the present study, the behavior expected of an effective performer was improved quantity of performance, yet it could have been some other aspect of behavior. Similarly, the role of effective performer could relate to that of a supervisor by specifying improve-

ment in planning and coordinating activities. In this way, the effective performer role can relate to a number of more general roles.

As the discovery of boundary conditions in this study underlines, certain conditions must be met in the design of a work organization before the predictions from this theory apply. These conditions have been outlined above and they should be recalled again. In addition to boundary conditions, another difference between the interdependent role systems theory and other contemporary approaches is the orientation toward dynamic as opposed to static correlations.

Dynamic Correlations

Interdependent role systems theory deals with changes in the work behavior of an individual over time. The criteria of interest, using this model, are changes in the individual's behavior relative to his past performance. Hence, this model requires the measurement of dynamic as opposed to static variation. The dimension of time is our most faithful ally. Although static correlations (the correlations between prediction variables and the level of performance at one time period for a group of persons) may be appropriate for variables commonly assumed to be stable over time (e.g., abilities), static correlations are not desirable for dynamic variables (e.g., motivational variables), because they are insensitive to changes in the behavior of individuals over time—the behavior of interest. In contrast, dynamic correlations (the correlations between prediction variables and the gain in performance over an interval of time for a group of persons) maximize the behavior of interest. The results of the present study supply ample evidence of this disguising effect of static correlations: Static correlations failed to support the hypothesis, although dynamic correlations did support the hypothesis concerning job performance.

Hypothesis Concerning Job Performance

Without further introduction the hypothesis derived from interdependent role systems theory for predicting the gain in performance

is as follows:

$$B = [(\sum_{i=1}^I A_i I_i) E'] w_0 + (\sum_{j=1}^J R_j P_j) w_1 + (\sum_{k=1}^K A_k E_k) w_2$$

($i = 1, 2, \dots, I; j = 1, 2, \dots, J;$
 $k = 1, 2, \dots, K$)

where

B = gain in performance.

A_i = preference for outcome i (attraction).

I_i = belief that the attainment of the work role of effective performer will lead to outcome i (instrumentality).

E' = the difference between the subjective probability that the act involving superior effort will lead to more effective performance and that for the act involving standard effort. ($E_1 - E_2$).

R_j = belief as to what person j expects him to do or not do (received role).

P_j = perceived pressure to comply with the expectations of person j .

A_k = preference for the intrinsic consequence k of the act (attraction).

E_k = expectancy that the act will lead to consequence k (expectancy).

w_0, w_1, w_2 = beta weights of a linear, multiple regression equation that may take any values.

The first term in this equation, the "path-goal utility" (Georgeopoulos, Mahoney, & Jones, 1957), is the attitude toward the act as a means to attain the role of effective performer with its accruing role outcomes. This term was represented by our effective performer model in the present study. Therefore, the above equation includes our effective performer model and two additional terms. These additional terms refer to the external and internal pressures on the individual to perform the act. The second term in the equation, "external pressure," is the individual's perceptions of what others expect him to do and the

pressure he feels they would apply to influence his compliance to their expectations. The third and final term in the equation, "internal pressure," is the individual's perceptions of the probability of various intrinsic consequences of the act and his preferences for attaining these various consequences.

In short, an individual's gain in the performance of a specific act in a given situation is a function of the following:

- 1a. His preferences for various role outcomes (attractions).
- 1b. His beliefs that the attainment of the work role of effective performer will lead to these various role outcomes (instrumentalities).
- 1c. His belief that increased effort in the performance of the act will lead to more effective performance (expectancy).
- 2a. His perceptions of what other persons expect him to do (received role).
- 2b. His perception of the amount of pressure these other persons would apply to influence his compliance (perceived pressure).
- 3a. His preferences for various intrinsic consequences of the act (attractions of consequences).
- 3b. His beliefs that the act will lead to these various intrinsic consequences (expectancy of consequences).

The determinants of the gain in behavior assumed by this theory thus include three classes of variables: (a) path-goal utility, (b) external pressure, and (c) internal pressure. Moreover, the relative importance of these three classes of variables depends on the particular act and the particular situation. The form of the equation that combines these three components is specified as linear, multiple regression; however, this specification is intended to be responsive to research on decision making.

Although the results of the present study supported the hypothesis about the relationship between path-goal utility (the effective performer model) and the gain in performance, it should be noted that path-goal utility is merely one component of this larger model. The second component, external pressure, in-

cludes considerations of the direction and magnitude of influence from various sources external to the individual. Each attempt to exert pressure on the individual includes two separate aspects: (a) information about what he should or should not do and (b) influence to elicit compliance with this information. The information aspect of this external pressure includes at least the two parameters of sign and relevance. The sign of information is either prescriptive or proscriptive. In addition, the influence aspect has at least the two parameters of the strength of the influence attempt and the perceived power of the person attempting to influence. The strength of the influence attempt is the individual's perception of the magnitude of pressure intended. In contrast, perceived power is the perception of the influencer's ability to control various consequences of compliance or noncompliance, such as gratifications, deprivations, and punishments. Therefore, external pressure includes a variety of psychological costs and rewards not included under path-goal utility.

The final component of this model, internal pressure, considers additional costs and rewards. This component includes two aspects: (a) the attraction of various intrinsic consequences of performing the act and (b) the expectancy that the act will lead to these various consequences. Some examples of favorable consequences of an act are satisfactions associated with performing the act (e.g., expressing valued abilities) and the satisfactions

associated with doing the task the way it should be done (e.g., complying with personal role expectations). In contrast, some examples of unfavorable consequences are fatigue, frustration, threats to both physical and psychological health and well-being, opportunity costs, and possible cognitive inconsistencies. As shown in Figure 3, the probability of superior effort is a function of the resolution of pressures toward and against superior effort, applied through path-goal utility, external sources, and internal sources.

Hypothesis Concerning Job Satisfaction

Work role satisfaction within this framework is determined by a more complex set of variables than have been assumed by past research. All three components of the job performance equations are employed to predict the gain in satisfaction. This hypothesis for predicting the gain in satisfaction is as follows:

$$S_n = \left[\left(\sum_{i=1}^I A_i I_i \right) E^* \right] b_0 + \left[\frac{\sum_{j=1}^{J'} R_j P_j}{\sum_{j=1}^J R_j P_j} \right] b_1 + \left[\frac{\sum_{k=1}^{K'} A_k E_k}{\sum_{k=1}^K A_k E_k} \right] b_2$$

$(i = 1, 2, \dots, I; j = 1, 2, \dots, J', \dots, J; k = 1, 2, \dots, K', \dots, K)$

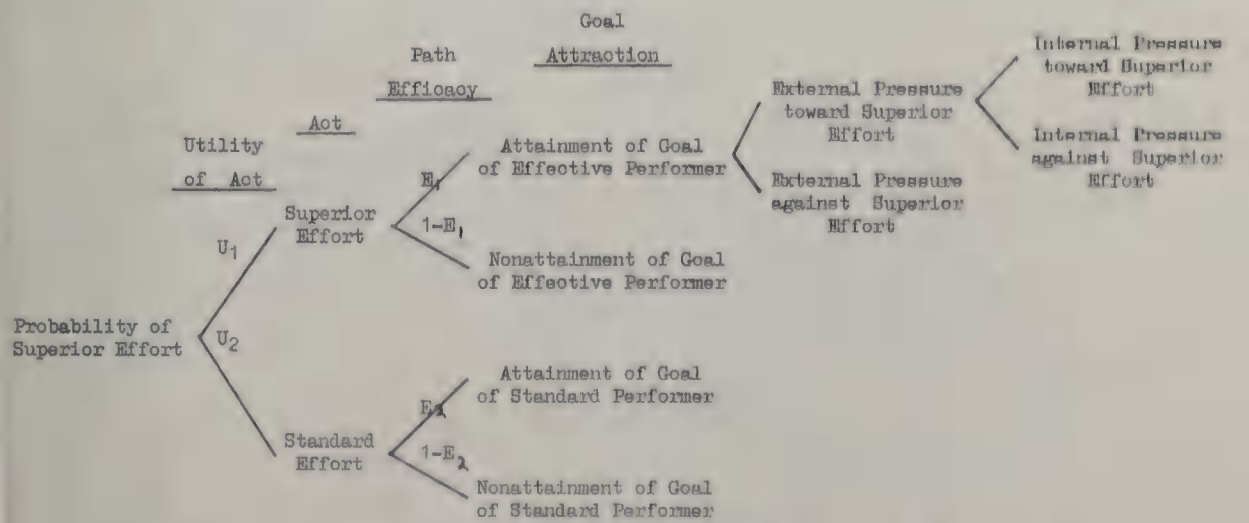


FIG. 3. Model from interdependent role systems theory for predicting the probability of superior effort.

where

- Sn = gain in satisfaction.
 $j = 1, 2, \dots, J'$ = external pressure consonant with the act.
 $k = 1, 2, \dots, K'$ = internal pressure consonant with the act.

This equation states that considering the most critical acts of a position, the gain in satisfaction is a function of the degree of path-goal utility, the amount of external pressure that is consonant with the given act relative to the total external pressure (both consonant and dissonant), and the amount of internal pressure that is consonant with the act relative to the total internal pressure. As shown in Figure 4, work role satisfaction is a function of the degree of path-goal utility and the extent to which both external and internal pressures can be resolved through appropriate behaviors.

These hypotheses predict changes in effort and satisfaction respectively as a consequence of being rewarded contingent upon effective performance by an organization that maintains a reciprocating climate. Space limitations do not permit a more complete statement of this theory, however, several suggested avenues for future research can be offered. The first component of the model, path-goal utility, should develop from research on the structure and formation of attitudes and expectancies. The second and

third components of the model should benefit from research on such concepts as power and authority, informational and normative influence, role expectations, and role conflict resolution. In addition, many of the ideas of Katz and Kahn (1966) about the process of role taking should be refined, stated as hypotheses, and tested.

Suggestions for Future Research

The results of this study imply that the work personality-work role systems approach, concerned with determining the effects of work role treatments on differing work personalities, is as promising as the approach concerned only with the average behavior effects of work role treatments, if not more promising. This system approach attempts to capitalize on individual differences in work personality by making predictions based upon work personality-work role interactions. Results further indicate that at least the two roles of job incumbent and effective performer make contributions to overall job satisfaction and the latter contributes to job performance. This suggests that it should prove worthwhile to explore other work roles, such as that of occupational development, within the systems framework. In addition, the analysis of organizational work groups from the work personality-work role systems approach should provide useful information on existing contingencies and interactions, knowledge presently lack-

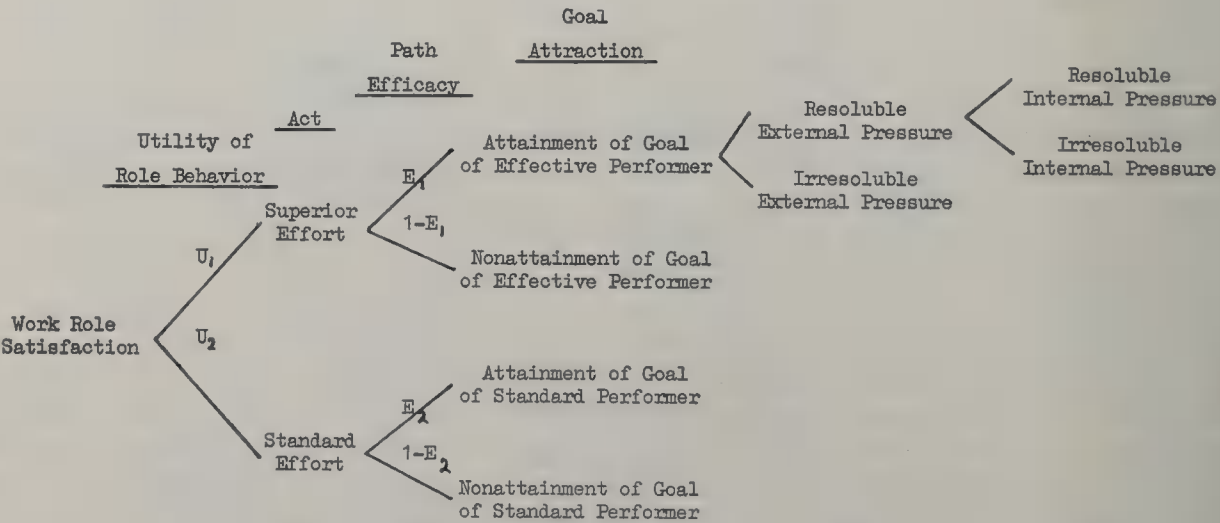


FIG. 4. Model from interdependent role systems theory for predicting work role satisfaction.

ing but needed for the effective employment of human talent. The need for a vast amount of additional work on the models of instrumentality and interdependent role systems theory is apparent. At the present time, instrumentality and interdependent role systems theories are best characterized as outlines that specify certain kinds of variables that should be important in understanding work motivation. The task of writing the text from either outline must rest upon the symbiotic interaction of researchers and theorists.

REFERENCES

- ADAMS, J. S. Toward an understanding of inequity. *Journal of Abnormal and Social Psychology*, 1963, 67, 422-436.
- ANDERSON, L. R., & FISHBEIN, M. A. Prediction of attitude from the number, strength, and evaluative aspect of beliefs about the attitude object: A comparison of summation and congruity theories. *Journal of Personality and Social Psychology*, 1965, 3, 437-443.
- ATKINSON, J. W. Towards experimental analysis of human motivation in terms of motives, expectancies, and incentives. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.
- CARLSON, E. R. Attitude change through modification of attitude structure. *Journal of Abnormal and Social Psychology*, 1956, 52, 256-261.
- DULANY, D. E. Hypotheses and habits in verbal "operant conditioning." *Journal of Abnormal and Social Psychology*, 1961, 63, 251-263.
- EWEN, R. B., SMITH, P. C., HULIN, C. L., & LOCKE, E. A. An empirical test of the Herzberg two-factor theory. *Journal of Applied Psychology*, 1966, 50, 544-550.
- FISHBEIN, M. A. Attitude and the prediction of behavior. In M. A. Fishbein (Ed.), *Readings in attitude theory and measurement*. New York: Wiley, 1967.
- GEORGOPOULOS, B. S., MAHONEY, G. M., & JONES, N. W. A path-goal approach to productivity. *Journal of Applied Psychology*, 1957, 41, 345-353.
- GALBRAITH, J., & CUMMINGS, L. L. An empirical investigation of the motivational determinants of task performance: Interactive effects between instrumentality-valence and motivation-ability. *Organizational Behavior and Human Performance*, 1967, 2, 237-257.
- GRAEN, G. B. Addendum to "An empirical test of the Herzberg two-factor theory." *Journal of Applied Psychology*, 1966, 50, 551-555.
- GRAEN, G. B. Work motivation: The behavioral effects of job-content and job-context factors in an employment situation. Unpublished doctoral dissertation, University of Minnesota, 1967.
- GRAEN, G. B. Testing traditional and two-factor hypotheses concerning job satisfaction. *Journal of Applied Psychology*, in press.
- GRAEN, G. B., & HULIN, C. L. Addendum to "An empirical investigation of two implications of the two-factor theory of job satisfaction." *Journal of Applied Psychology*, in press.
- GUILFORD, J. P. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.
- HARRIS, C. W. (Ed.) *Problems in measuring change*. Madison: University of Wisconsin Press, 1963.
- HOPPOCK, R. *Job satisfaction*. New York: Harper and Brothers, 1935.
- KATZ, D. N., & KAEN, R. L. *The social psychology of organizations*. New York: Wiley, 1966.
- PEAK, H. Attitude and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: University of Nebraska Press, 1955.
- PEAK, H. The effect of aroused motivation on attitudes. *Journal of Abnormal and Social Psychology*, 1960, 61, 463-468.
- Rosenberg, M. J. Cognitive structure and attitudinal affect. *Journal of Abnormal and Social Psychology*, 1956, 53, 367-372.
- ROSENTHAL, R. The effect of the experimenter on the results of psychological research. In B. Maher (Ed.), *Progress in Experimental Personality Research*, New York: Academic Press, 1964.
- ROTTER, J. B. The role of the psychological situation in determining the direction of human behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: University of Nebraska, 1955.
- TOLMAN, E. C. Principles of purposive behavior. In S. Koch (Ed.), Vol. 2. *Psychology: A study of science*. New York: McGraw-Hill, 1959.
- VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.
- WEICK, K. E. Organizations in the laboratory. In V. H. Vroom (Ed.), *Methods of organizational research*. Pittsburgh: University of Pittsburgh Press, 1967.
- WERNIMONT, P. F. Intrinsic and extrinsic factors in job satisfaction. *Journal of Applied Psychology*, 1966, 50, 41-50.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received March 11, 1968)

journal of applied psychology
monograph

Vol. 53, No. 3, Part 2

June 1969

"Test-Wiseness" on Personality Scales

By

Lawrence J. Stricker

Editor: KENNETH E. CLARK

Published by the American Psychological Association

journal of applied psychology

monograph

Vol. 53, No. 3, Part 2

June 1969

"TEST-WISENESS" ON PERSONALITY SCALES¹

LAWRENCE J. STRICKER²

Educational Testing Service

Test-wiseness on self-report personality scales was explored, using measures of accuracy in estimating the frequency of endorsement of personality items, estimating their social desirability, and identifying and "keying" items that measured the same factor, as well as indexes of ability to change scores on standard personality scales when they were administered with fake-good and fake-bad instructions. These variables generally did not correlate with each other, and they had only moderate and scattered correlations with personality scales administered with standard instructions. The test-wiseness measures were generally uncorrelated with ability and cognitive style tests and defensiveness scales, but they did correlate consistently with social desirability response style scales.

According to test lore, people vary in their knowledge about tests, and this "test-wiseness" affects their performance on these devices—the more test-wise obtaining higher scores on ability and aptitude tests and distorting their scores on personality inventories (Anastasi, 1961; Cronbach, 1960; Ebel & Damrin, 1960; Fishman, Deutsch, Kogan, North, & Whiteman, 1964; Goslin, 1963; Guilford, 1959; Pauck, 1950; Thorndike, 1949; Vernon, 1958, 1962). Despite the prevalence of these notions, the relevant data are sparse. Millman, Bishop, and Ebel (1965)

have reviewed the scattered findings for ability tests and described some aspects of test-wiseness on such measuring instruments. Even less is known about this phenomenon on personality inventories.

Test-wiseness on inventory measures of personality may involve several abilities. Perhaps the most complex is the ability to respond in accordance with a prescribed role in completing a personality questionnaire. Responding in this way probably reflects, in addition to its own particular form of test-wiseness, the presence of other abilities involved in test-wiseness and sheer knowledge of the role (Bordin, 1943; Gough, 1947). This particular kind of "impression management" (Goffman, 1959) is displayed in role-playing studies (Dahlstrom & Welsh, 1960; Ellis, 1953; Waters, 1965), which compare scores on personality scales administered with standard instructions with scores on the same scales administered with instructions to fake either a good role (e.g., a superbly well-adjusted person) or a bad one (e.g., a severely disturbed individual). People vary markedly

¹ This study was supported by the National Institute of Mental Health under Research Grant 1 P01 HD 01762. Portions of this study were presented at the meeting of the American Psychological Association, New York, September 1967. Thanks are due Anne Bloxom for locating and abstracting the studies reviewed in this article, Henrietta Gallagher for supervising the statistical analyses, and Bruce Bloxom and Fred L. Damarin for their critical reviews of a draft of this article.

² Requests for reprints should be sent to the author, Educational Testing Service, Princeton, New Jersey 08540.

in their success in faking (Gough, 1952; Grayson & Olinger, 1957; Hedberg, 1962; Hunt, 1948; Kimber, 1947; Lanyon, 1967; Noll, 1951). These individual differences are not related to intelligence (Kelly, Miles, & Terman, 1936; Kimber, 1947; Noll, 1951) or age (Kelly et al., 1936; Noll, 1951), but women are more skillful at faking than men (Kimber, 1947; Noll, 1951) and well-adjusted people are generally better at it than maladjusted ones (Canter, 1963; Grayson & Olinger, 1957; Hunt, 1948; Lanyon, 1967). Widely different findings about the generality of this ability have been reported; scores on the same scales administered with different role-playing instructions were unrelated in one study (Kelly et al., 1936) and highly related in others (Hedberg, 1962; Rusmore, 1956).

Another potentially relevant ability is accuracy in estimating the desirability of the items found on personality scales. Edwards' (1957) social desirability paradigm suggests that people's responses on these scales depend on the items' desirability in their society, implying that accurate knowledge of desirability is needed for socially desirable responding and, more generally, for dissembling on personality inventories. The two studies relevant to desirability estimation lend only moderate support to such a conception of this ability. In the first (Wiggins, 1966), Ss' accuracy in estimating the average desirability ratings made by a sample of Ss like themselves correlated with skill in faking-good on MMPI (Hathaway & McKinley, 1951) clinical scales, but did not correlate with success in faking-good on MMPI scales measuring test-taking attitudes.³ In the second investigation (Edwards, 1965), individual Ss' personal judgments of social desirability corresponded closely to average social desirability ratings from another group of Ss. But the extent of correspondence between their judgments and the average ratings, which may be a rough indication of their accuracy in estimating desirability, was uncorrelated with their scores on Edwards' (1957) Social

Desirability (*SD*) scale, a measure of socially desirable responding.

A similar ability is accuracy in estimating the "communality" (Wiggins, 1962) or frequency of endorsement of personality items. The desirability of personality items and their communality are highly related; Edwards (1953) reported that the two correlated .87. The relevant findings for this variable are negative: Accuracy in communality estimation did not correlate with ability to fake good on any of the MMPI scales that were studied (Wiggins, 1966).

Accuracy in analyzing a personality scale and determining the nature of the traits that it is intended to measure may also be a pertinent ability. Individual differences in the transparency of personality scales have not been studied in the context of test-wiseness, though procedures to measure this kind of skill have been developed and used (Hofstee⁴; Seeman, 1952). A survey (Fiske, 1967) indicated that this skill may be fairly common: A substantial proportion of the general population was aware that brief versions of the personality inventories which they completed as part of the survey were intended to measure "personality" or "stability."

The present study was designed to investigate systematically the role of these test-wiseness abilities on personality scales, using specially developed measures of these skills. The study's specific purposes were to determine (a) the prevalence of these abilities; (b) their generality; (c) their relationships with performance on standard personality scales; and (d) their links with other ability, cognitive style, and personality variables that may also be implicated in performance on the standard personality scales or on the test-wiseness instruments.

METHOD

Subjects

The Ss, paid volunteers, were 92 undergraduate women at an eastern state university. The results were analyzed for the 91 Ss for whom usable data were available.

⁴ William K. B. Hofstee, personal communication, undated.

³ These results were obtained with an "absolute accuracy" score, which is roughly analogous to the scores used in Edwards' (1965) study as well as in the present one.

Procedure

All the measures were obtained at the same testing session. At the outset, Ss were told the following:

I'm _____ from Educational Testing Service in Princeton. We are conducting a study, sponsored by the United States Government, attempting to evaluate the usefulness of ability tests and personality questionnaires. You'll be taking a variety of tests and questionnaires today. . . . We also want to emphasize that these tests and questionnaires are intended solely for research purposes.

The Ss were also asked to put their names on the tests and questionnaires.

The instruments were administered in the following order: (a) an inventory containing scales from the Guilford-Zimmerman Temperament Survey (GZTS; Guilford & Zimmerman, 1949) and using the standard GZTS instructions; (b) another inventory containing SD response style and defensiveness scales and administered with standard instructions adapted from the California Psychological Inventory (CPI; Gough, 1957); (c) Advanced Vocabulary Test (French, Ekstrom, & Price, 1963), Mathematics Aptitude Test (French et al., 1963), Letter Sets Test (French et al., 1963), Estimation Questionnaire (Pettigrew, 1958), and Object Sorting Test (Clayton & Jackson, 1961)—all ability or cognitive style measures; (d) Ability to Identify Items, Estimating Communalities, and Estimating Desirability—three test-wiseness measures; and (e) the inventory containing the GZTS scales, readministered with fake-good instructions and then with fake-bad instructions.

Test-Wiseness Measures⁵

Estimating desirability. This instrument was similar in rationale to Wiggins' (1966) measure, and, to a lesser extent, to the one employed by Edwards (1965). It was constructed by selecting randomly 19 items from each of the five *Dy* scales (Jackson & Messick, 1961) of the MMPI, after eliminating items on which 105 male and 85 female undergraduates at Stanford University differed significantly (χ^2 corrected for continuity, $p < .05$) in their endorsement frequencies (Wiggins, 1959, 1964a). The *Dy* scales represent five levels of SD and their items overlap minimally with standard MMPI clinical scales. The

⁵ Estimating Desirability, Estimating Communalities, and Ability to Identify Items, as well as their scoring keys, and the complete fake-good and fake-bad instructions have been deposited with the National Auxiliary Publications Service. Order Document No. 00363 from National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 22 West 4th Street, New York, New York 10001. Remit in advance \$3.00 for photocopies or \$1.00 for microfiche and make checks payable to: Research and Microfilm Publication, Inc.

⁶ All the significance tests described in this article are two-tailed.

95 items were administered in random order with instructions "to judge how desirable the average college student would consider the behavior or opinion that is described by each statement." Judgments were made on a 9-point scale, ranging from Extremely Undesirable to Extremely Desirable. The score was the product-moment correlation, transformed to Fisher's z , between S's desirability judgments for the 95 items and the items' SD scale values, as obtained by Messick and Jackson (1961) from data for 171 male and female undergraduates at Pennsylvania State University.⁷

Estimating communalities.⁸ This device paralleled the Estimating Desirability measure in design and resembled Wiggins' (1966) measure in its general rationale. It consisted of 19 items randomly selected from each of the *Dy* scales, after eliminating items with sex differences in endorsement frequencies and items used on the Estimating Desirability measure. The items were administered in random order with instructions "to judge how frequently college students would respond 'true' to each statement when answering a questionnaire describing themselves." Judgments were made on a 9-point scale that ranged from Extremely Infrequent to Extremely Frequent. The score was the transformed correlation between S's frequency judgments and the items' actual endorsement frequency for the Stanford undergraduates, combining both sexes.⁹

Ability to identify items.¹⁰ This test was adapted from procedures employed by Seeman (1952) and Hofstee (see Footnote 4). It consisted of three similarly constructed subtests. Each subtest was based on a different published factor analysis of personality items that obtained eight or more rotated and interpretable factors, including a factor loaded ($> .30$) by at least 8 items. The studies used were by Comrey and Soufi (1960, 1961) and Layman (1940). A subtest consisted of 15 items: the 8 with the highest loadings on the same factor and 7 others—each with the highest loading ($> .30$) on one of seven other

⁷ The corresponding correlation ratio (η), based on the regression of the items' SD scale values on S's desirability judgments, was also computed. Twenty-eight of the 91 correlation ratios were significantly ($p < .05$) greater than the corresponding product-moment correlation coefficients, indicative of nonlinearity in the regression, but the product-moment correlation between the two alternative kinds of scores on this instrument— η and transformed r —was .96.

⁸ This instrument was administered to Ss with the title, "Estimating Frequency."

⁹ The correlation ratio was also computed, based on the regression of the items' actual endorsement frequencies on S's frequency judgments. Eleven of the 91 correlation ratios were significantly greater than their respective correlation coefficients; the correlation between the two kinds of scores on this device was .97.

¹⁰ The title, "Ability to Identify Personality Characteristics," was used in administering this instrument to Ss.

factors. In choosing the main factor loaded by the 8 items in each analysis, an attempt was made to obtain factors that would be as different as possible in the three analyses. Those employed were "gregariousness" (Layman, 1940—Subtest A in this instrument), "cheerfulness vs. depression" (Comrey & Soufi, 1961—Subtest B), and "poor physical health" (Comrey & Soufi, 1960—Subtest C). In the Layman analysis the items were revised so that they were in the first person singular, corresponding to the wording of the items in the two other analyses; and in selecting items for the main factor one item was excluded because it was a direct reversal of an item with a higher loading which had already been chosen for that factor.

The 15 items in a subtest were presented in random order, with instructions to identify the items that refer to the same personality trait and to "key" them by indicating the response (True or False) that reflected the presence of the trait.

The subtest score was the sum of (a) the number of items from the main factor that were identified as involving the same trait and correctly keyed and (b) the number of items not from that factor which *S* indicated did not refer to the same trait. Since the main factors were bipolar, *S* would be equally correct in keying the items to correspond to either pole. Consequently, *Ss'* answers were scored twice, first keying them for one pole, then keying them for the opposite pole; *S* received the higher of the two scores. A total score for the instrument was obtained, weighting the subtest scores for optimal reliability (Green, 1950).

Role-playing measures. These procedures were similar to those used in most role-playing studies. Four GZTS scales—General Activity, Sociability, Emotional Stability, and Personal Relations—were chosen for this purpose because they were highly reliable, moderately intercorrelated, roughly balanced in the proportion of items keyed yes and no, and judged likely to shift under role-playing instructions. The scales were administered with fake-good and fake-bad instructions adapted from Yonge and Heist (1965). The fake-good instructions were

Imagine that you are in the following situation: You have applied for admission to a college or university. As part of the selection process, you have been asked to complete this questionnaire. Since you want to be accepted for admission, you wish to make the most *favorable* impression possible. Try to answer the questionnaire in a way that will make such a favorable impression on the admissions committee.

The fake-bad instructions were

Imagine that you are in the following situation: You have been forced to apply for admission to a certain university at the insistence of your parents. As part of the selection process, you have been asked to complete this questionnaire. Since you disagree with your parents' decision and do not want to be accepted for admission to this university, you wish to make the most *unfavorable* im-

pression possible. Try to answer the questionnaire in a way that will make such an unfavorable impression on the admissions committee.

Three scores were obtained for each scale: (a) the content score, using the published key for the scale, obtained with fake-good instructions; (b) the content score with fake-bad instructions; and (c) a difference score (McNemar, 1958, p. 48), representing the estimated "true" difference between these two scores (i.e., the fake-bad score minus the fake-good score). It was anticipated that the difference score would be the most sensitive role-playing measure, reflecting the actual influence of the role-playing instructions on *Ss'* responses. The fake-good and fake-bad scores, used by themselves, provide no base line for assessing the extent to which the same scores would have been achieved with other role-playing instructions or with standard instructions. Although the difference score necessarily is highly related to the fake-good and fake-bad scores from which it is derived, the latter were also employed in this study's analyses in order to provide continuity with previous investigations that used such scores.¹¹

Ability and Cognitive Style Measures

Measures of two major factors in the ability domain—verbal comprehension and general reasoning—were administered in order to determine their similarity to the abilities tapped by the test-wiseness measures. Verbal comprehension was measured by the Advanced Vocabulary Test—V4 (French et al., 1963) and general reasoning by the Mathematics Aptitude Test—R2 (French et al., 1963).

Other ability tests and the cognitive style measures were administered because of their potential relevance to performance on Ability to Identify Items. One such ability was induction, which might be involved in examining the items and picking out a common subset of them. The measure of this variable was the Letter Sets Test—Part 1 (French et al., 1963). The cognitive style of category width (Bruner, Goodnow, & Austin, 1956; Pettigrew, 1958) also seemed pertinent, for a predilection for overly wide categories could result in the judgment that the personality trait common to the items was very broad, encompassing all of them; a bias in favor of overly narrow categories would result in the opposite effect. Category width was measured by the Estimation Questionnaire (Pettigrew, 1958). Scores were obtained for Factor 1, involving time and speed items, and Factor 2, encompassing more general content. Similar reasons dictated the inclusion of another cognitive style,

¹¹ Role-playing ability could also be assessed by profile analysis procedures, which have the unique advantage, in principle, of detecting *Ss* with implausibly extreme scores or suspicious score patterns. These methods were not used in this study because they have only been fully developed for the MMPI (Dahlstrom & Welsh, 1960), necessarily requiring that the investigation focus on that inventory and involving the unduly time-consuming administration and readministration of the entire MMPI.

equivalence range (Gardner, 1953). A tendency toward broad equivalence range could result in the judgment that all the items were similar, and, hence, reflect the same trait; a preference for a narrow range would imply that none of the items had any similarity. A group-administered version (Clayton & Jackson, 1961) of the Object Sorting Test (Gardner, 1953) was used to measure equivalence range. Two scores were obtained from this test (Form I): the number of categories containing two or more objects and the number of miscellaneous objects left uncategorized (Messick & Kogan, 1963).

In view of the general relevance of psychological knowledge to test-wiseness, Ss were asked to list all the college-level courses in psychology that they had taken or were currently taking. The total number of credit hours was tabulated.¹²

SD and Defensiveness Measures

Validity scales on the MMPI and CPI as well as SD scales were included in the battery in view of their pertinence to personality inventory performance and in order to assess their similarity to the variables tapped by the test-wiseness measures. The particular scales that were chosen had been widely used in previous studies of this kind or represented distinctly different assessment strategies. Factor analyses (Edwards, 1963; Edwards, Diers, & Walker, 1962; Edwards & Walsh, 1964; Martin, 1964; Quinn & Lichtenstein, 1965; Wiggins, 1964b) indicate that these and other such measures fall roughly into two

groups: those that assess SD response style—Edwards' (1957) SD scale, Stricker's (1963) SD scale, the Wb scale (Gough, 1957), the K scale (Hathaway & McKinley, 1951), the F scale (Hathaway & McKinley, 1951), and the F—K index (Gough, 1950)—and those that gauge defensiveness or lying—Wiggins' (1959) Sd scale, the Marlowe-Crowne (Crowne & Marlowe, 1960) SD scale, and the L scale (Hathaway & McKinley, 1951).

RESULTS¹³

Means and Theoretical Score Limits of Test-Wiseness Measures

The means and standard deviations of the test-wiseness measures are reported in Table 1, together with the theoretical minimum and maximum scores on each measure.

A comparison of the mean scores with their theoretical limits indicates that Ss obtained high scores in absolute terms. The means for Estimating Desirability and Estimating Communality were 1.08 and .68, respectively, corresponding to mean correlations of .79 and .59. The means for Ability to Identify Items and for the role-playing measures were even closer to their theoretical limits. The mean was 15.86 for Ability to Identify Items, and

¹² Such variables as empathy and social intelligence may also be pertinent to test-wiseness, but they could not be included in this study because adequate measures of these characteristics are not available.

¹³ Tables reporting the means, standard deviations, and intercorrelations of all variables have been deposited with the National Auxiliary Publications Service. See Footnote 5 for ordering information.

TABLE 1
MEANS, STANDARD DEVIATIONS, AND THEORETICAL SCORE LIMITS OF
TEST-WISENESS MEASURES

| Measure | <i>M</i> | <i>SD</i> | Theoretical score limits | |
|--------------------------------------|----------|-----------|--------------------------|-----------------------|
| | | | Minimum test-wiseness | Maximum test-wiseness |
| Estimating Desirability | 1.08 | .16 | —3.00 | 3.00 |
| Estimating Communality | .68 | .27 | —3.00 | 3.00 |
| Ability to Identify Items | 15.86 | 1.20 | .00 | 19.39 |
| Fake-Good—General Activity | 21.84 | 3.35 | .00 | 30.00 |
| Fake-Good—Sociability | 26.82 | 2.40 | .00 | 30.00 |
| Fake-Good—Emotional Stability | 26.56 | 2.81 | .00 | 30.00 |
| Fake-Good—Personal Relations | 23.73 | 3.99 | .00 | 30.00 |
| Fake-Bad—General Activity | 5.70 | 4.12 | 30.00 | .00 |
| Fake-Bad—Sociability | 3.38 | 4.73 | 30.00 | .00 |
| Fake-Bad—Emotional Stability | 3.51 | 3.21 | 30.00 | .00 |
| Fake-Bad—Personal Relations | 2.86 | 2.33 | 30.00 | .00 |
| Difference Score—General Activity | —10.55 | 5.72 | 28.48 | —21.93 |
| Difference Score—Sociability | —13.35 | 5.71 | 30.18 | —18.74 |
| Difference Score—Emotional Stability | —18.56 | 4.31 | 25.77 | —24.37 |
| Difference Score—Personal Relations | —17.33 | 4.13 | 21.47 | —24.50 |

TABLE 2
INTERCORRELATIONS OF TEST-WISENESS MEASURES

| Measure | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) |
|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| (1) Estimating Desirability | (.63) | .23* | .07 | .05 | -.06 | .06 | .12 | -.06 | -.03 | .05 | .01 | -.02 | -.04 | .06 | .10 |
| (2) Estimating Communalities | | (.86) | .09 | .09 | -.02 | .06 | .14 | .15 | .05 | .08 | .07 | .14 | .03 | .08 | .14 |
| (3) Ability to Identify Items | | | (.62) | -.07 | .00 | -.09 | -.09 | .07 | .05 | .05 | -.08 | .02 | .04 | -.02 | -.10 |
| (4) Fake-Good—General Activity | | | | (.69) | .18 | .32** | .22* | .60** | .30** | .29** | .13 | .84** | .30** | .36** | .23* |
| (5) Fake-Good—Sociability | | | | | (.63) | .29** | .14 | .30** | .53** | .14 | .07 | .28** | .71** | .25* | .14 |
| (6) Fake-Good—Emotional Stability | | | | | | (.74) | .42** | .23* | .25* | .46** | .33** | .29** | .29** | .82** | .46** |
| (7) Fake-Good—Personal Relations | | | | | | | (.76) | -.09 | .12 | .19 | .33** | .03 | .13 | .34** | .92** |
| (8) Fake-Bad—General Activity | | | | | | | | (.82) | .53** | .43** | .11 | .94** | .52** | .39** | -.03 |
| (9) Fake-Bad—Sociability | | | | | | | | | (.90) | .40** | .13 | .49** | .98** | .39** | .14 |
| (10) Fake-Bad—Emotional Stability | | | | | | | | | | (.78) | .27** | .42** | .37** | .88** | .26** |
| (11) Fake-Bad—Personal Relations | | | | | | | | | | | (.61) | .13 | .12 | .35** | .67** |
| (12) Difference Score—General Activity | | | | | | | | | | | | (.86) | .49** | .42** | .07 |
| (13) Difference Score—Sociability | | | | | | | | | | | | | (.90) | .39** | .15 |
| (14) Difference Score—Emotional Stability | | | | | | | | | | | | | | (.84) | .40** |
| (15) Difference Score—Personal Relations | | | | | | | | | | | | | | | (.79) |

Note.—Internal-consistency reliability coefficients appear in parentheses.

* $p < .05$.

** $p < .01$.

the theoretical limits on this instrument were 0 to 19.39. The means for the fake-good scales ranged from 21.84 to 26.82, those for the fake-bad scales ranged from 2.86 to 5.70, and the limits on both kinds of scales were 0 to 30.00. The means for the difference scores ranged from -10.55 to -18.56; their limits ranged from -18.74 to -24.50 on one side, and from 21.47 to 30.18 on the other.

The fake-good means were consistently higher than the corresponding fake-bad means. These differences were highly significant ($p < .001$)—the t ratios, computed for dependent groups, were 22.90 for General Activity, 35.03 for Sociability, 42.53 for Emotional Stability, and 37.68 for Personal Relations.

Intercorrelations of Test-Wiseness Measures

The product-moment intercorrelations of the test-wiseness measures appear in Table 2 together with estimates of their internal-consistency reliability. Reliability was estimated by the Spearman-Brown formula from the correlation between split-halves for Estimating Desirability and Estimating Communality, by Green's (1950) procedure for Ability to Identify Items, by Cronbach's (1951) Coefficient Alpha for the fake-good and fake-bad scales, and by Lord's (1956) Formula 27 for the difference scores. In this analysis, as well as in subsequent ones, the correlations with the fake-bad scales and difference scores have been reflected in sign so that, in effect, high scores on all test-wiseness measures represent high ability.

Except for the .23 correlation between Estimating Desirability and Estimating Communality, none of the correlations between the four kinds of instruments—Estimating Desirability, Estimating Communality, Ability to Identify Items, and the role-playing devices—was significant ($p > .05$). The role-playing measures generally correlated significantly with each other, but the correlations were moderate. Their median intercorrelations were .26 for the fake-good scales, .34 for the fake-bad scales, and .40 for the difference scores. The corresponding fake-good and fake-bad scales consistently correlated with each other—.60 for General Activity, .53 for Sociability, .46 for Emotional Stability, and .33 for Personal Relations—and fake-good and

fake-bad versions of different scales were frequently correlated too.

Correlations of Test-Wiseness Measures with Standard GZTS Scales

Total group. The product-moment correlations of the test-wiseness measures with the four GZTS scales administered with standard instructions appear in Table 3 for the total group of Ss. The Coefficient Alpha reliability estimates for the four scales are also reported in this table.

The test-wiseness measures had some significant but moderate correlations with the GZTS scales. Apart from a .20 correlation between Estimating Communality and the Personal Relations scale, all the correlations of Estimating Desirability, Estimating Communality, and Ability to Identify Items were limited to the Emotional Stability scale (r 's were .21, .27, and -.29, respectively). The role-playing measures generally correlated with their own GZTS scale, when it was administered with standard instructions—the consistent exception was the Sociability scale—but they had few correlations with other GZTS scales. The fake-good scales that correlated with the corresponding standard scales were General Activity ($r = .31$), Emotional Stability ($r = .32$), and Personal Relations ($r = .52$); the fake-bad scales that correlated were General Activity ($r = .20$) and Emotional Stability ($r = .30$); and the difference scores involved were General Activity ($r = .27$), Sociability ($r = .36$), and Personal Relations ($r = .47$). The role-playing measures that correlated with standard scales other than their own were the fake-good Personal Relations scale ($r = .30$ with the Emotional Stability scale), the fake-bad Emotional Stability scale ($r = .20$ with the Sociability scale), and the Personal Relations difference score ($r = .27$ with the Emotional Stability scale).

Subgroups defined by response styles. The product-moment correlations of the test-wiseness measures with the standard GZTS scales, computed separately for Ss above and below the median on composite measures of SD response style and defensiveness, appear in Table 4.¹⁴

¹⁴ The composite measure of SD response style was the sum of the standard scores on Edwards' SD

TABLE 3

CORRELATIONS OF TEST-WISENESS MEASURES WITH STANDARD GZTS SCALES

| Measure | Standard GZTS Scale | | | |
|--------------------------------------|---------------------|-------------|---------------------|--------------------|
| | General Activity | Sociability | Emotional Stability | Personal Relations |
| Estimating Desirability | .08 | .02 | .21* | .11 |
| Estimating Communality | .06 | -.07 | .27** | .20* |
| Ability to Identify Items | .16 | .04 | -.29** | -.04 |
| Fake-Good—General Activity | .31** | .07 | .04 | -.01 |
| Fake-Good—Sociability | .04 | -.03 | -.09 | -.10 |
| Fake-Good—Emotional Stability | .16 | .03 | .32** | .12 |
| Fake-Good—Personal Relations | .04 | -.10 | .30** | .52** |
| Fake-Bad—General Activity | .20* | .08 | -.10 | -.15 |
| Fake-Bad—Sociability | -.08 | .00 | .04 | -.02 |
| Fake-Bad—Emotional Stability | .15 | .20* | .30** | .15 |
| Fake-Bad—Personal Relations | .14 | -.02 | .08 | .14 |
| Difference Score—General Activity | .27** | .09 | -.05 | -.11 |
| Difference Score—Sociability | -.06 | -.01 | .01 | -.05 |
| Difference Score—Emotional Stability | .18 | .14 | .36** | .16 |
| Difference Score—Personal Relations | .08 | -.09 | .27** | .47** |
| Internal-Consistency Reliability | .84 | .85 | .85 | .80 |

* $p < .05$.** $p < .01$.

There were scattered differences in the correlations for the high and low groups. In 5 of the 60 pairs of correlations in the SD response style analysis, the correlations for the two groups were significantly different ($p < .05$, using a z test for transformed correlations): Estimating Desirability correlated $-.16$ with the Emotional Stability scale in the high SD response style group and $.40$ with this scale in the low group; the fake-good General Activity scale correlated $.55$ and $.08$ with the standard version of this scale in the two groups, the fake-good Emotional Stability scale correlated $.24$ and $-.20$ with the Sociability scale, the fake-bad Personal Relations scale correlated $.40$ and $-.04$ with the standard version of this scale, and the Emotional Stability difference score cor-

related $.38$ and $-.11$ with the Sociability scale. In the defensiveness analysis, 4 pairs of correlations were significantly different: Estimating Communality correlated $.20$ with the Sociability scale in the high defensiveness group and $-.27$ in the low one, the fake-good Sociability scale correlated $.17$ and $-.28$ with the Emotional Stability scale in the two groups, the fake-bad Sociability scale correlated $-.28$ and $.17$ with the General Activity scale, and the Emotional Stability difference score correlated $.37$ and $-.04$ with the Personal Relations scale.

Correlations of Test-Wisness Measures with Ability, Cognitive Style, SD, and Defensiveness Measures

scale, Stricker's *SD* scale, the *Wb* scale, the *K* scale, and the *F* scale, weighting each score equally. Since a high score on the *F* scale, unlike the other scales, reflects low social desirability, its scores were reversed before summing. The corresponding composite measure of defensiveness was computed in the same way, using Wiggins' *Sd* scale, the Marlowe-Crowne *SD* scale, and the *L* scale. The Coefficient Alpha reliability was $.91$ for the first composite measure and $.72$ for the second.

The product-moment correlations of the test-wisness measures with the ability and cognitive style tests and the SD response style and defensiveness measures appear in Table 5. Estimates of the internal-consistency reliability of most of these measures are also reported. Reliability was estimated by the Spearman-Brown formula for all the ability tests except the Letter Sets Test; by the correlation between Parts 1 and 2 for the

TABLE 4
CORRELATIONS OF TEST-WISENESS MEASURES WITH STANDARD GZTS SCALES FOR SUBGROUPS DEFINED BY RESPONSE STYLE MEASURES

| Standard GZTS Scale | | | | | | | | | | | |
|---------------------------------------|------------------|--------------|-------|--------------|--------------|-------|---------------------|--------------|--------|--------------------|--------------|
| Measure | General Activity | | | Sociability | | | Emotional Stability | | | Personal Relations | |
| | Above median | Below median | z | Above median | Below median | z | Above median | Below median | z | Above median | Below median |
| SD response style groups ^a | | | | | | | | | | | |
| Estimating Desirability | .07 | .07 | .02 | -.14 | .05 | .87 | -.16 | .40** | 2.71** | .03 | .12 |
| Estimating Communality | .09 | -.04 | .56 | -.14 | -.19 | .21 | .07 | .25 | .88 | .14 | .13 |
| Ability to Identify Items | .29* | .07 | 1.08 | .14 | .11 | .13 | -.11 | -.34* | 1.14 | .21 | -.16 |
| Fake-Good—General Activity | .55** | .08 | 2.49* | .25 | -.06 | 1.46 | -.04 | .12 | .74 | -.07 | .03 |
| Fake-Good—Sociability | .26 | -.15 | 1.95 | .04 | -.09 | .61 | -.08 | -.17 | .44 | -.15 | -.11 |
| Fake-Good—Emotional Stability | .25 | .06 | .93 | .24 | -.20 | 2.05* | .28 | .26 | .11 | .03 | .07 |
| Fake-Good—Personal Relations | .04 | -.01 | .26 | -.02 | -.26 | 1.15 | .14 | .28* | .71 | .51** | .48** |
| Fake-Bad—General Activity | .30* | .09 | 1.03 | .17 | .03 | .66 | -.11 | -.09 | .06 | -.11 | -.20 |
| Fake-Bad—Sociability | -.01 | -.18 | .80 | .03 | -.07 | .47 | -.06 | .05 | .53 | .08 | -.18 |
| Fake-Bad—Emotional Stability | .19 | .09 | .48 | .38* | .01 | 1.83 | .25 | .39** | .76 | .16 | .11 |
| Fake-Bad—Personal Relations | .04 | .16 | .55 | -.15 | -.07 | .40 | .01 | -.04 | .23 | .40** | -.04 |
| Difference Score—General Activity | .42** | .10 | 1.59 | .21 | -.01 | 1.05 | -.09 | .00 | .40 | -.10 | -.12 |
| Difference Score—Sociability | .06 | -.18 | 1.10 | .04 | -.08 | .54 | -.08 | -.01 | .31 | .04 | -.17 |
| Difference Score—Emotional Stability | .25 | .08 | .80 | .38** | -.11 | 2.38* | .31* | .38** | .36 | .13 | .11 |
| Difference Score—Personal Relations | .05 | .06 | .06 | -.07 | -.24 | .82 | .11 | .21 | .48 | .53** | .36** |

Table 4—(Continued)

| Measure | Standard GZTS Scale | | | | | | | | | | |
|--------------------------------------|---------------------|--------------|----------|--------------|--------------|----------|---------------------|--------------|----------|--------------------|--------------|
| | General Activity | | | Sociability | | | Emotional Stability | | | Personal Relations | |
| | Above median | Below median | <i>z</i> | Above median | Below median | <i>z</i> | Above median | Below median | <i>z</i> | Above median | Below median |
| Defensiveness groups ^b | | | | | | | | | | | |
| Estimating Desirability | .15 | .04 | .52 | .09 | .09 | .02 | .24 | .31* | .36 | .18 | .14 |
| Estimating Communality | .09 | .03 | .30 | .20 | -.27 | 2.21* | .38** | .19 | .94 | .24 | .19 |
| Ability to Identify Items | .14 | .23 | .42 | .00 | .21 | .97 | -.26 | -.26 | .04 | .06 | -.12 |
| Fake-Good—General Activity | .49** | .18 | 1.62 | .28 | -.02 | 1.45 | .19 | -.05 | 1.14 | .00 | .00 |
| Fake-Good—Sociability | .10 | -.01 | .51 | .16 | -.14 | 1.40 | .17 | -.28* | 2.12* | -.03 | -.17 |
| Fake-Good—Emotional Stability | .29* | .05 | 1.13 | .05 | -.10 | .71 | .38** | .23 | .76 | .29* | -.09 |
| Fake-Good—Personal Relations | .22 | -.13 | 1.62 | .04 | -.21 | 1.18 | .42** | .21 | 1.06 | .61** | -.45** |
| Fake-Bad—General Activity | .13 | .27 | .67 | .08 | .10 | .08 | -.07 | -.13 | .27 | -.11 | -.19 |
| Fake-Bad—Sociability | -.28 | .17 | 2.10* | .00 | .06 | .27 | .14 | -.03 | .76 | .06 | -.09 |
| Fake-Bad—Emotional Stability | -.05 | .25 | 1.38 | .18 | .15 | .13 | .46** | .19 | 1.39 | .33* | .01 |
| Fake-Bad—Personal Relations | .13 | .14 | .04 | -.18 | .01 | .91 | .05 | .07 | .08 | .32* | .00 |
| Difference Score—General Activity | .30* | .25 | .25 | .18 | .06 | .59 | .04 | -.11 | .67 | -.08 | -.13 |
| Difference Score—Sociability | -.22 | .13 | 1.65 | .04 | .01 | .14 | .16 | -.11 | 1.23 | .04 | -.12 |
| Difference Score—Emotional Stability | .14 | .19 | .22 | .14 | .05 | .41 | .50** | .24 | 1.39 | .37** | -.04 |
| Difference Score—Personal Relations | .21 | -.04 | 1.16 | -.02 | -.16 | .63 | .34* | .20 | .71 | .58** | .36** |

^a Forty-four Ss were in the group above the median and 47 in the group below it.
^b Forty-five Ss were in the group above the median and 46 in the group below it.
* $p < .05$.
** $p < .01$.

TABLE 5
CORRELATIONS OF TEST-WISENESS MEASURES WITH ABILITY, COGNITIVE STYLE, AND RESPONSE STYLE MEASURES

| Measure | Adv. Vocab. | Math. Apt. | Ltr. Sets | Est. Quest. | | Object Sorting | | Psych. credit hours | Edw. SD | Str. SD | Wb | K | F | F-K | Wig. SD | M-C SD | L |
|--------------------------------------|----------------|---------------|--------------|-------------|-------------|----------------|--------|---------------------------|------------|------------|-------|------|--------|--------|------------|-----------|-------|
| | | | | Factor 1 | Factor 2 | Categ. | Misc. | | | | | | | | | | |
| Estimating Desirability | .11 | .07 | -.10 | .04 | .08 | .09 | .06 | .03 | .26** | .14 | .23* | .12 | -.27** | -.22* | .02 | .06 | -.04 |
| Estimating Communality | .10 | .08 | .12 | -.17 | .05 | -.09 | .09 | .25* | .26** | .24* | .20* | .23* | -.31** | -.31** | -.11 | .06 | -.04 |
| Ability to Identify Items | .02 | .09 | .14 | -.22* | -.06 | .01 | .05 | .04 | -.22* | -.20* | -.23* | -.16 | .20* | .21* | -.04 | -.12 | -.20* |
| Fake-Good—General Activity | .17 | .11 | .02 | .21* | .17 | -.15 | .09 | -.05 | .06 | -.11 | -.02 | -.10 | .11 | .12 | .00 | -.08 | .10 |
| Fake-Good—Sociability | .00 | -.16 | -.16 | -.01 | -.14 | .10 | -.27** | -.03 | -.15 | -.08 | -.01 | -.08 | .03 | .07 | -.08 | .01 | -.09 |
| Fake-Good—Emotional Stability | .13 | -.04 | .06 | .16 | .15 | -.18 | -.13 | .11 | .24* | .16 | .23* | .09 | -.19 | -.16 | .08 | .20* | .11 |
| Fake-Good—Personal Relations | .02 | .02 | .08 | .09 | .13 | .05 | -.16 | .17 | .23* | .24* | .06 | .13 | -.13 | -.16 | .02 | .03 | .04 |
| Fake-Bad—General Activity | .16 | .06 | .02 | .02 | -.01 | -.19 | -.04 | .06 | -.05 | -.15 | -.02 | -.06 | .12 | .10 | -.01 | -.13 | -.05 |
| Fake-Bad—Sociability | .12 | -.10 | -.07 | -.03 | -.05 | .10 | -.24* | .06 | .01 | .03 | .05 | .00 | .01 | .00 | -.25* | -.15 | -.11 |
| Fake-Bad—Emotional Stability | .18 | .05 | .15 | .08 | .10 | -.33** | .04 | .06 | .26** | .18 | .20* | .15 | -.13 | -.17 | .11 | .12 | .18 |
| Fake-Bad—Personal Relations | .14 | .18 | -.01 | -.03 | .09 | -.11 | .09 | -.11 | .20* | .08 | .05 | -.03 | -.24* | -.10 | .10 | -.03 | .01 |
| Difference Score—General Activity | .18 | .08 | .02 | .11 | .07 | -.19 | .01 | .02 | .00 | -.15 | -.02 | -.08 | .13 | .12 | -.01 | -.12 | .01 |
| Difference Score—Sociability | .10 | -.13 | -.10 | -.03 | -.08 | .11 | -.27** | .04 | -.03 | .00 | .05 | -.02 | .01 | .02 | -.23* | -.12 | -.12 |
| Difference Score—Emotional Stability | .19 | .01 | .13 | .14 | .15 | -.31** | -.04 | .10 | .29** | .20* | .25* | .15 | -.18 | -.19 | .12 | .18 | .17 |
| Difference Score—Personal Relations | .07 | .08 | .06 | .06 | .14 | -.01 | -.09 | .09 | .26** | .22* | .07 | .09 | -.20* | -.16 | .05 | .01 | .03 |
| Internal-Consistency Reliability | .78 | .72 | .48 | .81 | .81 | — | — | — | .82 | .73 | .81 | .70 | .59 | .75 | .35 | .83 | .47 |

* $p < .05$.
** $p < .01$.

latter test¹⁵; and by Coefficient Alpha for the Estimation Questionnaire as well as for the response style and defensiveness scales.

The test-wiseness measures did not correlate significantly with any ability test, and they had only scattered correlations with the cognitive style measures, mostly involving the Object Sorting Test. The fake-bad Emotional Stability scale correlated ($r = -.33$) with the Categories score on the Object Sorting Test, and so did the difference score on this scale ($r = -.31$). The fake-good, fake-bad, and difference measures of Sociability all correlated with the Miscellaneous Objects score on the Object Sorting Test (r 's were $-.27$, $-.24$, and $-.27$, respectively). Ability to Identify Items and the fake-good General Activity scale both correlated with the Factor 1 score on the Category Width Test (r 's were $-.22$ for the former measure and $.21$ for the latter). And Estimating Communality correlated $.25$ with psychology credit hours.

Estimating Desirability, Estimating Communality, and Ability to Identify Items consistently had significant correlations with the SD measures; the two accuracy instruments were positively related to socially desirable responding, but Ability to Identify Items was negatively related to this response style. All correlated with Edwards' SD scale (r 's were $.26$, $.26$, and $-.22$, respectively), the *Wb* scale (r 's were $.23$, $.20$, and $-.23$), the *F* scale (r 's were $-.27$, $-.31$, and $.20$), and the *F - K* index (r 's were $-.22$, $-.31$, and $.21$). Both Estimating Communality and Ability to Identify Items correlated with Stricker's SD scale, $.24$ for the former, $-.20$ for the latter. Estimating Communality also correlated $.23$ with the *K* scale.

In contrast to the wide correlations of these test-wiseness instruments with the SD measures, only one correlated with any defensiveness scale—Ability to Identify Items correlated $-.20$ with the *L* scale.

Virtually all the significant correlations for the role-playing measures involved the Emotional Stability and Personal Relations scales, on the one hand, and the SD scales, on the other; role-playing success on these GZTS

scales was positively related to socially desirable responding. The fake-good and fake-bad Emotional Stability scales and their difference score correlated $.24$, $.26$, and $.29$, respectively, with Edwards' SD scale; the corresponding correlations with the *Wb* scale were $.23$, $.20$, and $.25$. The difference score for Emotional Stability also correlated $.20$ with Stricker's SD scale. The pattern of correlations for the Personal Relations measures was similar to the one for the Emotional Stability measures. All three measures correlated with Edwards' SD scale (r 's were $.23$, $.20$, and $.26$, respectively). In addition, the fake-good version correlated $.24$ with Stricker's SD scale, and the difference score correlated $.22$ with this SD scale. The fake-bad scale and the difference score correlated $-.24$ and $-.20$, respectively, with the *F* scale.

Like the other test-wiseness instruments, the role-playing measures had few correlations with the defensiveness scales. The fake-bad Sociability scale and the difference score on this scale both correlated with Wiggins' *Sd* scale (r 's were $-.25$ for the fake-bad scale and $-.23$ for the difference score); and the fake-good Emotional Stability scale correlated $.20$ with the Marlowe-Crowne SD scale.

Correlations of SD and Defensiveness Measures with Standard GZTS Scales

The product-moment correlations of the SD response style and defensiveness measures with the GZTS scales administered with standard instructions appear in Table 6.

The Sociability, Emotional Stability, and Personal Relations scales correlated significantly with all the SD measures; high scores on these GZTS scales were positively related to socially desirable responding. The three correlated with Edwards' SD scale (r 's were $.46$, $.75$, and $.43$), Stricker's SD scale (r 's were $.38$, $.66$, and $.55$), the *Wb* scale (r 's were $.43$, $.62$, and $.41$), the *K* scale (r 's were $.42$, $.58$, and $.44$), the *F* scale (r 's were $-.25$, $-.44$, and $-.38$), and the *F - K* index (r 's were $-.42$, $-.62$, and $-.49$). The General Activity scale correlated with only one of these response style measures—Edwards' SD scale ($r = .22$).

¹⁵ This correlation was based on a subgroup of 45 Ss with scores on both parts of this test.

TABLE 6

CORRELATIONS OF RESPONSE STYLE MEASURES WITH STANDARD GZTS SCALES

| GZTS Scale | Edw. SD | Str. SD | Wb | K | F | F-K | Wig. Sd | M-C SD | L |
|---------------------|------------|------------|-------|-------|--------|--------|------------|-----------|-------|
| General Activity | .22* | .05 | .15 | .03 | -.02 | -.03 | .27** | .07 | -.01 |
| Sociability | .46** | .38** | .43** | .42** | -.25* | -.42** | .35** | .31** | .25* |
| Emotional Stability | .75** | .66** | .62** | .58** | -.44** | -.62** | .19 | .42** | .24* |
| Personal Relations | .43** | .55** | .41** | .44** | -.38** | -.49** | .06 | .29** | .26** |

* $p < .05$.
 ** $p < .01$.

The GZTS scales also correlated with the defensiveness scales—high scores on the GZTS scales were positively related to defensiveness, but the correlations were lower in level and less consistent: The General Activity and Sociability scales correlated with Wiggins' *Sd* scale (r 's were .27 and .35); and the Sociability, Emotional Stability, and Personal Relations scales correlated with both the Marlowe-Crowne *SD* scale (r 's were .31, .42, and .29) and the *L* scale (r 's were .25, .24, and .26).

DISCUSSION

Prevalence of Test-Wiseness

The high level of test-wiseness displayed in this study was striking. These results are consistent with those obtained previously. Marked shifts in scores have been routinely observed in role-playing studies (Dahlstrom & Welsh, 1960; Ellis, 1953; Waters, 1965), indicating considerable skill in this activity. And, in Edwards' (1965) investigation, *Ss*' desirability ratings predicted *SD* scale values with great accuracy. In the face of the considerable test-wiseness that exists, it is noteworthy that wide and reliable individual differences were observed in this study, as well as in earlier ones.

In evaluating the present findings, it should be noted that *Ss*, prior to the study, probably resembled most college students in the extent of their exposure to personality inventories and to other experiences that might produce test-wiseness. Although high already, their test-wiseness, like other abilities, could probably be improved by appropriate training (Ebel & Damrin, 1960). The relationship between amount of course work in psychology

and Estimating Communality, though modest, suggests that the ability tapped by this instrument, at least, is amenable to training.

Generality of Test-Wiseness

Another important finding was the specificity of the test-wiseness measures: The different kinds of instruments were unrelated, with one minor exception, though there was some generality among the role-playing variables.¹⁶ This finding implies that test-wiseness is not a broad, general ability, but consists of a set of distinct and largely unrelated skills. The confirmation of this inference requires studies of other populations of *Ss*, especially more naive ones who are known to have no experience with personality inventories, and by appraisals of the test-wiseness measures' construct validity. The logical possibility exists that the abilities tapped by the measures simply did not reflect test-wiseness or only sampled limited aspects of it. Indeed, the particular abilities that were studied probably do not exhaust those in the test-wiseness domain; still, they do seem to be a good representation of this domain, with the possible exception of the variable tapped by Ability to Identify Items, which functioned in unexpected ways. The operation of this mea-

¹⁶ Some of the relationships among the role-playing variables may be due, in part, to shared variance introduced by using the same instructions or item content for two or more variables. In contrast, all the relationships with and among the other test-wiseness measures were heterotrait, heteromethod (Campbell & Fiske, 1959), for these instruments differed among themselves and from the role-playing variables in both instructions and item content. (Different, but essentially equivalent, sets of items were used in Estimating Desirability and Estimating Communality.)

sure will be discussed in more detail subsequently.

Especially puzzling was the role-playing measures' complete independence of Estimating Desirability and Estimating Communality. The results obtained by Wiggins (1966), in a similar investigation, had generally been negative, too. The present study and Wiggins' study point to the conclusion that accuracy in estimating desirability and communality is not an important ingredient in role-playing success. A minimal degree of accuracy may be essential, but this is probably exceeded by most Ss, judging from the extent of test-wiseness that has been observed. Very special skills may be involved in projecting oneself into a role. The role in this study, for example, had two distinct requirements that had to be mastered, and these demanded a combination of test-wiseness and role knowledge. One was assuming the guise of an applicant for college admission. This necessitated accurate perception of the general range of responses appropriate for college-age high school graduates, as contrasted, for example, with those associated with institutionalized psychiatric patients. It is likely that Ss varied in the accuracy of these perceptions. The second was determining the kind of performance on the inventory that would impress or repel the admissions committee. Just as they would in a real life counterpart of this situation, Ss taking this role undoubtedly differed in their views of what would impress or repel the committee and in their ability to respond in a way that would achieve the effect that they desired. For these reasons the responses elicited by such role-playing instructions are apt to be considerably more complex than, and sharply divergent from, judgments of social desirability or communality. One indication of such a divergence is that role-playing responses typically vary for different roles that are similar in overall favorability (Hedberg, 1962; Krug, 1958; Norman, 1963; Rusmore, 1956; Wesman, 1952; Yonge & Heist, 1965), indicating that role-playing is not based on a generalized notion of social desirability or communality. In relatively uncomplex role-playing situations, the responses that are elicited may have a greater resemblance to judgments of social desira-

bility and communality—the simplest role-playing instructions explicitly or implicitly solicit these judgments. In such situations, some correspondence, perhaps a great deal, would be expected between role-playing success and accuracy in estimating these item characteristics.

Test-Wiseness and Performance on Standard Personality Scales

The Ss had the ability to distort their scores on personality scales, as gauged from the extent of test-wiseness that was observed, but the limited relationships between the test-wiseness measures and the GZTS personality scales imply that test-wiseness was not a major source of dissembling. It seems reasonable to assume that at least some dissembling occurred on the personality scales, and the strong possibility that the distortion was considerable is suggested by the scales' extensive correlations with the response style measures (Edwards, 1957; Jackson & Mesick, 1958), though these correlations may be open to less sinister, substantive interpretations, too (Block, 1965). The failure of the accuracy measures and Ability to Identify Items to correlate with scores on the fake-good and fake-bad scales suggests that the same kind of negative results would also be observed in real life situations where greater incentives to distort exist. This suggestion is extremely tenuous, however, because of the uncertain correspondence between such role-playing scores and those obtained in real life.

The general lack of relationship between the test-wiseness measures and the standard personality scales raises the obvious question: Why wasn't test-wiseness closely linked with distortion on these scales? At this juncture, this question cannot be answered with any certainty. One plausible answer is that test-wiseness did, indeed, produce distortion, but this relationship could not be uncovered in this study. The link between the two could have been obscured by the heterogeneity of Ss, who may have consisted of (a) those unmotivated to distort, (b) those motivated to distort in a favorable direction, and (c) those motivated to distort in an unfavorable direction—Ss in the last two groups varying in the extent of their motivation. However, the

general absence of differences between the correlations of the test-wiseness measure with the personality scales for those with high and those with low levels of SD response style and defensiveness argues against such an explanation, insofar as the response style measures reflect the extent of dissembling and the direction that it takes. These moderator-variable analyses are hampered, though, by the circularity introduced through the conceptual and empirical linkage between the response style and test-wiseness variables, and by the blurring of distinctions among the three possible kinds of Ss that was produced by dichotomizing the score distributions for the response style measure. Another possibility, as mentioned previously, is that the test-wiseness measures lacked construct validity. If, in fact, test-wiseness was not a major source of distortion, no compelling explanation can be offered for why this should be so. One speculation, difficult to verify, is that dissembling is largely unconscious, and, for this reason, not dependent on the skills involved in the test-wiseness measures.

In the face of the generally negative results that were obtained, the frequent, though moderate, correlations between the role-playing measures and their *own* standard personality scales provide an interesting exception. These correlations, when considered together with the overall lack of correlation of these same role-playing measures with *other* standard personality scales, furnish additional evidence for the specificity of test-wiseness that was already observed in the intercorrelation analysis. The present findings indicate that role-playing success on a particular scale, even when it was related to performance on the same scale administered with standard instructions, did not generalize to performance on other standard personality scales.

Test-Wiseness and Other Variables

All of the different kinds of test-wiseness measures, including Estimating Desirability, correlated extensively with the SD measures, though Edwards (1965) had found that a rough measure of accuracy in desirability estimation was uncorrelated with an SD scale. These extensive correlations contrast sharply with the scanty correlations of the test-wise-

ness measures with the defensiveness scales. The explanation may lie in differences in the item composition of the two kinds of response style scales (Jackson & Messick, 1962). The SD scales contain items with comparable social desirability and communality—the desirable items are frequently endorsed and the undesirable items are rarely endorsed. The items on the defensiveness scales typically have discrepant desirability and communality—the desirable items are seldom endorsed (e.g., "I read in the Bible several times a week.") and the undesirable ones are often endorsed (e.g., "At times I feel like swearing."). Such defensiveness items are deviant because of the high relationship between social desirability and communality, and are atypical of most items found on personality inventories, including the GZTS scales, as well as on Estimating Desirability, Estimating Communality, and Ability to Identify Items. The complex and unusual interaction of desirability and communality on the defensiveness items, their rareness, and their underrepresentation on the test-wiseness measures should attenuate the correlations of the defensiveness scales with the test-wiseness measures. A different explanation of these findings is that the defensiveness scales may tap a characteristic that is under less conscious control than the trait measured by the SD scales. Socially desirable responding may entail a relatively systematic consideration of the items' desirability and communality, accounting for its link with the accuracy measures; defensiveness, because it may not involve such considerations, is unrelated to knowledge of these item characteristics.

The finding that the various kinds of test-wiseness measures correlated with the SD scales, but, in general, did not correlate with each other suggests that different processes may underlie their relationships with this response style. One conjecture is that the link with Estimating Desirability and Estimating Communality, two measures that were related, may be causal, as indicated in the preceding discussion: Accurate knowledge of the desirability and communality of personality items is a prerequisite for choosing and making desirable and common responses. In contrast, the relationship between this response

style and role-playing may occur because they are symptomatic of the same personality dynamics.

The tie between the test-wiseness measures and SD response style may also explain why most of the test-wiseness measures were related to the Emotional Stability scale, but unrelated to the other standard personality scales. The Emotional Stability scale also correlated with the SD measures; indeed, it was more highly correlated than any of the other GZTS scales with all the SD measures. As a result, the relationship between the test-wiseness measures and this personality scale may have been produced by the response style variance common to them. In fact, eliminating, by partial-correlation procedures, the SD response style variance shared by the test-wiseness measures and this personality scale markedly attenuated the measures' correlations with the scale.¹⁷

The general lack of relationship between the test-wiseness devices and the ability tests and cognitive style measures in this study suggests that the test-wiseness skills exist independently of these other variables. The generality of this inference needs to be established by studies of other groups of Ss. This finding is consistent with previous reports that role-playing success and intelligence were uncorrelated (Kelly et al., 1936; Kimber, 1947; Noll, 1951). Although it would be useful to investigate the relationship of the test-wiseness measures to a broader spectrum of such variables, the abilities and cognitive styles

already studied were specifically chosen for their manifest relevance to test-wiseness, reducing the likelihood that further work along this line will uncover links between test-wiseness and other abilities or cognitive styles.

Functioning of Ability to Identify Items

The deviant pattern of correlations for Ability to Identify Items (e.g., its negative correlations with the standard Emotional Stability scale and the SD measures in contrast with the positive or nonsignificant correlations of the other test-wiseness measures with these variables) suggests that the ability it measures is not an element of test-wiseness, at least in the way that this phenomenon was originally conceived. This instrument's correlations with the other variables may have simply occurred for reasons unrelated to test-wiseness (e.g., its negative correlations with the Emotional Stability scale and the SD measures could be produced because the analytic people who do well on this instrument may also be introspective and overly sensitive to their own personality defects).

Overview

In considering the larger meaning of this study, the main theme that emerges is the failure to implicate test-wiseness as an important source of distortion on personality scales. Pending subsequent confirmation, it appears that test-wiseness, despite its prevalence, is not a functional unity and is not extensively linked to performance on inventory measures of personality. Additionally, test-wiseness seems to be unrelated to other potentially relevant variables, apart from its interesting tie with SD response style. These results raise a variety of provocative issues about the causes of distortion and the processes associated with test-wiseness. From a broader perspective, the findings underscore how little is currently known about the determinants of responses on self-report devices.

REFERENCES

- ANASTASI, A. *Psychological testing*. (2nd ed.) New York: Macmillan, 1961.
BLOCK, J. *The challenge of response sets*. New York: Appleton-Century-Crofts, 1965.

¹⁷ In this analysis, composite measures of SD response style and defensiveness were separately partialled out of all of the correlations between the test-wiseness measures and the GZTS personality scales administered with standard instructions. In computing these partial correlations, the correlations with the control variable were corrected for attenuation in this variable to prevent the undercorrection of the other variables arising from the unreliability of the control variable (Kahneman, 1965).

The only noticeable difference between the original correlations and the partial correlations was that most of the significant correlations of the test-wiseness measures with the Emotional Stability scale were reduced to nonsignificant levels when SD response style was held constant. A table containing all the partial correlations has been deposited with the National Auxiliary Publications Service. See Footnote 5 for ordering information.

- BORDIN, E. S. A theory of vocational interests as dynamic phenomena. *Educational and Psychological Measurement*, 1943, 3, 49-65.
- BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.
- CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, 56, 81-105.
- CANTER, F. M. Simulation on the California Psychological Inventory and the adjustment of the simulator. *Journal of Consulting Psychology*, 1963, 27, 253-256.
- CLAYTON, M. B., & JACKSON, D. N. Equivalence range, acquiescence, and overgeneralization. *Educational and Psychological Measurement*, 1961, 21, 371-382.
- COMREY, A. L., & SOUFI, A. Further investigation of some factors found in MMPI items. *Educational and Psychological Measurement*, 1960, 20, 777-786.
- COMREY, A. L., & SOUFI, A. Attempted verification of certain personality factors. *Educational and Psychological Measurement*, 1961, 21, 113-127.
- CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, 16, 297-334.
- CRONBACH, L. J. *Essentials of psychological testing*. (2nd ed.) New York: Harper & Row, 1960.
- CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 1960, 24, 349-354.
- DAHLSTROM, W. G., & WELSH, G. S. *An MMPI handbook*. Minneapolis: University of Minnesota Press, 1960.
- EBEL, R. L., & DAMRIN, D. E. Tests and examinations. In C. W. Harris (Ed.), *Encyclopedia of educational research*. (3rd ed.) New York: Macmillan, 1960.
- EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, 1953, 37, 90-93.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- EDWARDS, A. L. A factor analysis of experimental social desirability and response set scales. *Journal of Applied Psychology*, 1963, 47, 308-316.
- EDWARDS, A. L. Measurement of individual differences in ratings of social desirability and in the tendency to give socially desirable responses. *Journal of Experimental Research in Personality*, 1965, 1, 91-98.
- EDWARDS, A. L., DIERS, C. J., & WALKER, J. N. Response sets and factor loadings on sixty-one personality scales. *Journal of Applied Psychology*, 1962, 46, 220-225.
- EDWARDS, A. L., & WALSH, J. A. Response sets in standard and experimental personality scales. *American Educational Research Journal*, 1964, 1, 52-61.
- ELLIS, A. Recent research with personality inventories. *Journal of Consulting Psychology*, 1953, 17, 45-49.
- FISHMAN, J. A., DEUTSCH, M., KOGAN, L., NORTH, R., & WHITEMAN, M. Guidelines for testing minority group children. *Journal of Social Issues*, 1964, 20, 129-145.
- FISKE, D. W. The subject reacts to tests. *American Psychologist*, 1967, 22, 287-296.
- FRENCH, J. W., EKSTROM, R. B., & PRICE, L. A. *Manual for Kit of Reference Tests for Cognitive Factors*. Princeton, N. J.: Educational Testing Service, 1963.
- GARDNER, R. W. Cognitive styles in categorizing behavior. *Journal of Personality*, 1953, 22, 214-233.
- GOFFMAN, E. *The presentation of self in everyday life*. Garden City, N. Y.: Doubleday Anchor Books, 1959.
- GOSLIN, D. A. *The search for ability*. New York: Russell Sage Foundation, 1963.
- GOUGH, H. G. Simulated patterns on the Minnesota Multiphasic Personality Inventory. *Journal of Abnormal and Social Psychology*, 1947, 42, 215-225.
- GOUGH, H. G. The F minus K dissimulation index for the Minnesota Multiphasic Personality Inventory. *Journal of Consulting Psychology*, 1950, 14, 408-413.
- GOUGH, H. G. On making a good impression. *Journal of Educational Research*, 1952, 46, 33-42.
- GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1957.
- GRAYSON, H. M., & OLINGER, L. B. Simulation of "normalcy" by psychiatric patients on the MMPI. *Journal of Consulting Psychology*, 1957, 21, 73-77.
- GREEN, B. F., JR. A note on the calculation of weights for maximum battery reliability. *Psychometrika*, 1950, 15, 57-61.
- GUILFORD, J. P. *Personality*. New York: McGraw-Hill, 1959.
- GUILFORD, J. P., & ZIMMERMAN, W. S. *Manual of instructions and interpretations, the Guilford-Zimmerman Temperament Survey*. Beverly Hills, Calif.: Sheridan, 1949.
- HATHAWAY, S. R., & MCKINLEY, J. C. *Manual, Minnesota Multiphasic Personality Inventory*. (Rev. ed.) New York: Psychological Corporation, 1951.
- HEDBERG, R. More on forced-choice test fakability. *Journal of Applied Psychology*, 1962, 46, 125-127.
- HUNT, H. F. The effect of deliberate deception on Minnesota Multiphasic Personality Inventory performance. *Journal of Consulting Psychology*, 1948, 12, 396-402.
- JACKSON, D. N., & MESSICK, S. Content and style in personality assessment. *Psychological Bulletin*, 1958, 55, 243-252.
- JACKSON, D. N., & MESSICK, S. Acquiescence and desirability as response determinants on the MMPI. *Educational and Psychological Measurement*, 1961, 21, 771-790.

- JACKSON, D. N., & MESSICK, S. Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology*, 1962, **65**, 285-299.
- KAHNEMAN, D. Control of spurious association and the reliability of the controlled variable. *Psychological Bulletin*, 1965, **64**, 326-329.
- KELLY, E. L., MILES, C. C., & TERMAN, L. M. Ability to influence one's score on a typical pencil-and-paper test of personality. *Character and Personality*, 1936, **4**, 206-215.
- KIMBER, J. A. M. The insight of college students into the items on a personality test. *Educational and Psychological Measurement*, 1947, **7**, 411-420.
- KRUG, R. E. The effect of specific selection sets on a forced-choice self-description inventory. *Journal of Applied Psychology*, 1958, **42**, 89-92.
- LANYON, R. I. Simulation of normal and psychopathic MMPI personality patterns. *Journal of Consulting Psychology*, 1967, **31**, 94-97.
- LAYMAN, E. M. An item analysis of the adjustment questionnaire. *Journal of Psychology*, 1940, **10**, 87-106.
- LORD, F. M. The measurement of growth. *Educational and Psychological Measurement*, 1956, **16**, 421-437.
- MARTIN, J. Acquiescence—measurement and theory. *British Journal of Social and Clinical Psychology*, 1964, **3**, 216-225.
- MCMENAR, Q. On growth measurement. *Educational and Psychological Measurement*, 1958, **18**, 47-55.
- MESSICK, S., & JACKSON, D. N. Desirability scale values and dispersions for MMPI items. *Psychological Reports*, 1961, **8**, 409-414.
- MESSICK, S., & KOGAN, N. Differentiation and compartmentalization in object-sorting measures of categorizing style. *Perceptual and Motor Skills*, 1963, **16**, 47-51.
- MILLMAN, J., BISHOP, C. H., & EBEL, R. An analysis of test-wiseness. *Educational and Psychological Measurement*, 1965, **25**, 707-726.
- NOLL, V. H. Simulation by college students of a prescribed pattern on a personality scale. *Educational and Psychological Measurement*, 1951, **11**, 478-488.
- NORMAN, W. T. Personality measurement, faking, and detection: An assessment method for use in personnel selection. *Journal of Applied Psychology*, 1963, **47**, 225-241.
- PAUCK, C. E. A square deal for the examination. *Education*, 1950, **71**, 222-225.
- PETTIGREW, T. F. The measurement and correlates of category width as a cognitive variable. *Journal of Personality*, 1958, **26**, 532-544.
- QUINN, R. P., & LICHTENSTEIN, E. Convergent and discriminant validities of acquiescence measures. *Journal of General Psychology*, 1965, **73**, 93-104.
- RUSMORE, J. T. Fakability of the Gordon Personal Profile. *Journal of Applied Psychology*, 1956, **40**, 175-177.
- SEEMAN, W. "Subtlety" in structured personality tests. *Journal of Consulting Psychology*, 1952, **16**, 278-283.
- STRICKER, L. J. Acquiescence and social desirability response styles, item characteristics, and conformity. *Psychological Reports*, 1963, **12**(2, Pt. 2).
- THORNDIKE, R. L. *Personnel selection*. New York: Wiley, 1949.
- VERNON, P. E. Educational testing and test-form factors. Research Bulletin 58-3. Princeton, N. J.: Educational Testing Service, 1958.
- VERNON, P. E. The determinants of reading comprehension. *Educational and Psychological Measurement*, 1962, **22**, 269-286.
- WATERS, L. K. A note on the "fakability" of forced-choice scales. *Personnel Psychology*, 1965, **18**, 187-191.
- WESMAN, A. G. Faking personality-test scores in a simulated employment situation. *Journal of Applied Psychology*, 1952, **36**, 112-113.
- WIGGINS, J. S. Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology*, 1959, **23**, 419-427.
- WIGGINS, J. S. Strategic, method, and stylistic variance in the MMPI. *Psychological Bulletin*, 1962, **59**, 224-242.
- WIGGINS, J. S. An MMPI item characteristic deck. *Educational and Psychological Measurement*, 1964, **24**, 137-141. (a)
- WIGGINS, J. S. Convergences among stylistic response measures from objective personality tests. *Educational and Psychological Measurement*, 1964, **24**, 551-562. (b)
- WIGGINS, J. S. Social desirability estimation and "faking good" well. *Educational and Psychological Measurement*, 1966, **26**, 329-341.
- YONGE, G. D., & HEIST, P. A. The influence of suggested contexts on faking a personality test. *American Educational Research Journal*, 1965, **2**, 137-144.

(Received August 12, 1968)

